DOCUMENT RESUME

ED 243 928	TM 840 237
AUTHOR	Garvue, Robert; And Others
TITLE	The 1983 National Teacher Examinations Core Battery Louisiana Validation Study: Final Report. (Revised).
INSTITUTION	Louisiana State Dept. of Education, Baton Rouge. Office of Research and Development.
PUB DATE	22 Jun 83
NOTE	191p.; Prepared by the Louisiana State Dept. of Education NTE Task Force and presented to the National Teacher Examinations Blue Ribbon Score Committee.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	*Cutting Scores; Fine Arts; Higher Education; Listening Skills; Literature; *Occupational Tests; Professional Training; Quantitative Tests; Reading Tests; Science Tests; Social Studies; *Teacher Certification; *Teacher Education Curriculum; Teacher Educators; Teacher Supply and Demand; Test Use; *Test Validity; Writing Evaluation
IDENTIFIERS	*Louisiana; *National Teacher Examinations; Test Content

ABSTRACT

The purpose of this study was to determine the validity of the National Teacher Examinations (NTE) Core Battery for use in teacher certification in Louisiana. This information was produced to assist in the recommendation and establishment of a required score on the NTE. The jury judgment approach was used. Panels of faculty members were drawn from Louisiana institutions of higher education to validate the content of the tests in the Core Battery and estimate the score that could be expected from a minimally knowledgeable teacher candidate. The Content Review Panel found that 93 percent of the Professional Knowledge Test, 98 percent of the Mathematics Test, 97 percent of the Science Test, 100 percent of the Social Studies Test, 84 percent of the Literature/Fine Arts Test, 100 percent of the Reading Test, 98 percent of the Listening Test, and 96 percent of the objective items on the Writing Test were content appropriate. The Panel concluded that the NTE Core Battery is a valid measure for the teacher education curricula in Louisiana. The Knowledge Estimation Panel evaluated the difficulty and importance of each item of each of the eight Core Battery Tests, then estimated the proportion of items that the minimally knowledgeable teacher candidate would answer correctly. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243928

Accountability
Development
Evaluation
Management Information Systems
Research

THE 1983 NATIONAL TEACHER EXAMINATIONS CORE BATTERY LOUISIANA VALIDATION STUDY: FINAL REPORT

JUNE 22, 1983
(revised)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L. M. Hoffman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

STATE OF LOUISIANA
DEPARTMENT OF EDUCATION
Office of Research and Development
P.O. Box 44064
Baton Rouge, Louisiana 70804

J. Kelly Nix, Superintendent

TM 840 237

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

This document was published at a cost of $8.75 per copy by the State Department of Education, P.O. Box 44064, Baton Rouge, LA to provide leadership for the continuous development, coordination, and improvement of education on a statewide basis under authority of R.S. 17:21. This material was printed in accordance with the standards for printing by state agencies established pursuant to R.S. 43:31.

June 22, 1983

THE 1983 NATIONAL TEACHER EXAMINATIONS CORE BATTERY
LOUISIANA VALIDATION STUDY: FINAL REPORT

Prepared by:

Robert Garvue (Task Force Chairman)

Carol Falkowski

Lee Hoffman

Michael McGuire

Craig Mills

Janella Rachal

David Ransen

Charles Teddlie

Prepared For:

Superintendent J. Kelly Nix

and

The National Teacher Examinations
Blue Ribbon Score Committee

ACKNOWLEDGMENTS

This study was a cooperative project that required the attention and efforts of many people. The Louisiana college and university faculty members who served as panelists on the Validation Team had a crucial role; they are listed at the end of this report, but their contributions should be recognized at the beginning. Educational Testing Service furnished information and was generous in sharing the study procedures it had developed in the past. Dr. Hugh Peck provided encouragement, support, and good advice throughout the study. Dr. Helen Brown, Mr. Robert Crew, Mr. Michael Glisson, Dr. Jacqueline Lewis, Mr. Frank Norris and Mr. Joseph Williams, Jr. of the Louisiana Department of Education also provided advice and acted as a helpful sounding board as the design and implementation of the study emerged.

Special thanks must go to the people who were most important in dealing with the hard work and difficult timelines involved in producing this report. These were Ms. Nancy Wagner, who edited the final copy of the report, and Ms. Betty Alexander, Ms. Claire Holliday and Ms. Pat Davis who prepared the text in all of its stages.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS i

CHAPTER I. INTRODUCTION AND OVERVIEW 1
  Purpose of the Study. 1
  Score Recommendation Process. 3
    State Superintendent of Education. 4
    Blue Ribbon Score Committee. 4
    Validation Team. 5
    Technical Support. 5
  Study Design and Rationale. 6
    Use of the NTE 7
    Design 9
  References Cited. 14

CHAPTER II. PANEL SELECTION. 15
  Panel Selection Process 15
  Panel Nominations 17
    Representativeness of Panels 21
  Panel Assignments 26
    Content Review Panel
      Assignments and Attendance 26
    Knowledge Estimation Panel
      Assignments and Attendance 28
  Half-Panels 30
    Selection and Composition of Half-Panels 31
  Summary 33

CHAPTER III. CONTENT REVIEW RESULTS. 35
  Introduction. 35
    Content Appropriateness of Items 35
    Comprehensiveness of Tests 39
    Relative Emphasis of Tests and Curricula 41
    Overall Similarity Between
      Core Battery Tests and Curricula 44
  Interpreting and Evaluating Content Review Results. 45
    Criteria For Special Review. 47
    Special Review Criteria Results. 50
    Application of Special Review Criteria Results 52
  Consistency of Results 55
    Consistency of Results for
      Content Appropriateness of Items. 55
    Consistency of Results for Relative Emphasis 61
  Summary 64

TABLE OF CONTENTS (cont.)

Page

CHAPTER IV. KNOWLEDGE ESTIMATION RESULTS 66
  Introduction. 66
  Introduction to Standard Setting. 66
    Nedelsky 68
    Angoff 69
    Ebel 70
    Description of the Ebel Methodology Employed 71
    Description of the Essay Methodology 73
  Knowledge Estimation Process. 74
    Purpose of Knowledge Estimation Panels 74
    Formulation of a Reference Group 75
    Formulation of the Estimates 77
    Knowledge Estimation Training Session. 78
    Description of Ratings 79
    Relationship of Panel Judgments
      to Module Qualifying Scores 81
  Knowledge Estimation Results. 82
    Essay Ratings. 82
    Ebel Results 84
    Calculation of Judges' Standards 85
  Consistency of Results for Knowledge Estimation Process 89
    Consistency of Ebel Results. 90
    Consistency of Essay Results 95
    Conclusions. 97
  References Cited. 98

CHAPTER V. INFORMATION ABOUT RISK AND DEMAND 99
  Format of Revised NTE Qualifying Score. 100
  Risk of Rejecting Qualified Candidates. 101
  Supply of New Teachers in Louisiana 110

CHAPTER VI. SUMMARY OF THE STUDY 113
  Study Design. 113
  Study Participants. 114
  Content Review Findings 115
    Professional Knowledge Module. 115
    General Knowledge Module 116
    Communication Skills Module. 118
  Knowledge Estimation. 119
  Summary 120

APPENDICES 122
  I-A. NTE Blue Ribbon Score Committee. 122
  I-B. NTE Content Review and
    Knowledge Estimation Panel Members. 126
  I-C. Louisiana State Department
    of Education NTE Task Force Members 131

Page

III-A. Overview of Content Review Panel Tasks 133
III-B. Question Review Form and Instructions. 136
III-C. Test Content Descriptions, Test Content Review Forms, and Instructions 139
III-D. Test Content Summary Form. 151
IV-A. Item Estimation Form 153
IV-B. Overview of Knowledge Estimation Panel Tasks 155
IV-C. Description of the Minimally Knowledgeable Teacher Candidate 158
IV-D. Instructions for Knowledge Estimation Forms and KNowledge Estimation Training Exercise. 160
IV-E. Test Estimation Form 165
IV-F. Knowledge Estimation Training Worksheet. 167

LIST OF TABLES

Page

Table 1. Number Of Faculty Members Requested (Req) And Nominated (Nom); Shown By Institution And Area 18
Table 2. Allocation of Faculty Members Selected, Shown By Institution And Test. 20
Table 3. Percentage Of Graduates And Faculty Nominated, Selected, And Participating, By Institution. 22
Table 4. Faculty Members Nominated, Selected, And Participating From Institutions Classified By Type Of Support And Predominant Racial Composition Of Student Body: Percentages 24
Table 5. Faculty Members Nominated, Selected, And Participating By Sex And Years Of Experience Teaching In Louisiana: Numbers And Percentages. 25
Table 6. Number Of Faculty Members Selected And Attending Content Review Panel Assembly, Shown By Test Judged And Institution 27
Table 7. Number Of Faculty Members Selected And Attending Knowledge Estimation Panel Assembly, Shown By Test Judged And Institution 29
Table 8. Number And Characteristics Of Judges On Content Review And Knowledge Estimation Half-Panels, By Test . . . 32
Table 9. Number And Percent Of Items Rated Appropriate For Each Test Form And Module And Average Percent Appropriate For Test Forms And Modules 38
Table 10. Topics In The Curriculum Identified By Judges As Not Included In The Test Content Description Of The NTE Core Battery 40
Table 11. Frequencies Of Judgments About Relative Emphasis In Tests And Curriculum. 42
Table 12. Frequencies Of Judgments About Similarity Between The Core Battery Tests And The Teacher Education Sequence. 46
Table 13. Criteria For Performing Special Review Of Correspondence Between Test Content And Program Content . . 48
Table 14. Special Review Criteria Data For Core Battery Tests 51
Table 15. Conclusions About Correspondence Between Core Battery Modules And Program Content 53
Table 16. Number And Percent Of Questions Judged Content Appropriate, By Test Form And Half-Panel. 56
Table 17. Agreement Between Half-Panels About Content Appropriateness Of Items, By Test Form. 58
Table 18. Average Probability Of Agreement On Item Appropriateness Across Half-Panels. 60
Table 19. Half-Panel Ratings For Relative Emphasis Of Tests And Curricula 63
Table 20. Proportion Of Louisiana Essays Receiving Each Possible Score, NTE Core Battery Adminstration, November, 1982 . . . 83
Table 21. Average Percent Of Essays Classified In Each Judgment Category By NTE Writing Panel Members. 83
Table 22. Median And Mean Of Scores Established By Judges As Representing Borderline Essays. 84

LIST OF TABLES (cont.)

Page

Table 23A. Mean Number And Proportion Of Item Judgments Placed In Each Relevance Category By Knowledge Estimation Panel Members 86

Table 23B. Number And Percent of Items Rated Essential Or Important By A Majority Of Judges for Each Test Form And The Average Percent Of The Two Test Forms. 87

Table 24. Mean Number And Proportion Of Item Judgments Placed In Each Difficulty Category By Knowledge Estimation Panel Members. 88

Table 25. Results Of Application Of Ebel Methodology Standards By NTE Core Battery Module 89

Table 26. Results Of Application Of Ebel Methodology Standards By NTE Core Battery Module, Form, And Half-Panel: Scale Scores 91

Table 27. Comparison Of Half-Panel Ratings On Item Difficulty By Test Form 92

Table 28. Comparison Of Half-Panel Ratings On Item Relevance By Test Form. 94

Table 29. Comparison Of Half-Panel Ratings On Percent Of Items The Minimally Knowledgeable Teacher Candidate Would Be Expected To Answer Correctly. 96

Table 30. Average Percent Of Essays Classified In Each Judgment Category By Half-Panel. 96

Table 31. Probabilities Of Incorrectly Rejecting Qualified Applicants And Incorrectly Accepting Unqualified Applicants At Six Levels Of The Standard Error Of Measurement. 103

Table 32. Mean Examinee Score And Percent Of Examinees Meeting Judges' Standards On NTE Core Battery Modules. . . 104

Table 33. Percentage Of Louisiana Examinees Scoring At Or Above Study Estimate Of Core Battery Score Standards By Area Examination. 105

Table 34. Percentage Of Louisiana Examinees Scoring At Or Above Study Estimate Of Core Battery Standards Minus One SEM By Area Examination. 106

Table 35. Percentage of Louisiana Examinees Scoring At Or Above Study Estimate Of Core Battery Score Standards Minus Two SEMs By Area Examination 107

Table 36. Percentage Of Louisiana Examinees Scoring At Or Above Study Estimate Of Core Battery Score Standards Minue Three SEMs By Area Examination 108

LIST OF FIGURES

Page

Figure 1. Validation Decision Flow 3

Figure 2. Study Design 11

Figure 3. Ebel Decision Matrix 72

CHAPTER I

INTRODUCTION AND OVERVIEW

PURPOSE OF THE STUDY

The purpose of the study reported here was to determine the validity of the National Teacher Examinations (NTE) Core Battery for use in teacher certification in Louisiana. This information was produced to assist in the recommendation and establishment of a required score on the NTE. The report is presented to the Blue Ribbon Score Committee for its consideration in recommending a score to the State Superintendent of Education, and to the Superintendent for his deliberation in receiving the Committee's recommendation and determining the score he will refer to the State Board of Elementary and Secondary Education.

Louisiana law requires that any person applying for initial certification as a teacher (including those certified to teach in another state and applying for certification in Louisiana) shall satisfactorily pass an examination that includes English proficiency, pedagogical knowledge, and knowledge of his or her area of specialization (R.S. 17:7(6)). The law also requires that the State Superintendent of Education choose the appropriate instrument, conduct research to validate the applicability of the instrument to teacher education programs in Louisiana, and carry out research to determine the level at which the selected test is satisfactorily completed. In conducting this research, the Superintendent shall meet with, and consider the suggestions of, classroom teachers, representatives from teacher organizations, deans of education from Louisiana public colleges and universities, and representatives of the governing boards for higher education.

Educational Testing Service completed a validation study of the NTE Common Examinations and Area Examinations for the Louisiana Department of Education in July, 1978. Following this study, the State Superintendent of Education a) determined which components of the NTE were valid for use in Louisiana and b) established scores required for certification in 18 teaching areas. The score required for certification in each area was a composite of the Weighted Common Examinations Total and the appropriate Area Examination.

Since that time, Educational Testing Service, the publisher of the NTE, has replaced the Common Examinations with the Core Battery. The first administration of the new Core Battery was in November, 1982. As the outline below illustrates, there are substantive differences between the Common Examinations considered in the 1978 validation study and the Core Battery that was the object of this current study.

Common Examinations	Core Battery
● one 195-minute test	● three 120-minute modules
● components in Professional Education and General Knowledge with objective measures of typical teacher education training and liberal arts basics	● separate modules in General Knowledge, Communication Skills, and Professional Knowledge emphasizing the teacher as a problem solver and decision maker
● content in professional education, written English expression, social studies, literature and the fine arts, science, and mathematics	● content in professional education, social studies, literature and the fine arts, science, mathematics, reading, listening, and writing (objective questions and essay)
● scores in four areas (Professional Education, Written English Expression, Social Studies and Literature and Fine Arts, and Science and Mathematics) weighted to produce a Weighted Common Examinations Total	● unweighted scores from each of the three modules (General Knowledge, Communication Skills, and Professional Knowledge)

In the judgment of the Louisiana Department of Education, the differences between the Common Examinations and Core Battery warranted a complete validation study of the Core Battery. The Department was directed by the Superintendent of Education to conduct a validation study to be submitted to the Superintendent and an appointed committee that would assess the validity of the Core Battery and assist in the establishment of a qualifying score on the NTE for teacher certification.

SCORE RECOMMENDATION PROCESS

Three major groups are involved in the validation of the Core Battery and the establishment of a new NTE score required for teacher certification in Louisiana. These are the Superintendent of Education, the Blue Ribbon Score Committee, and the Validation Teams. Figure 1 outlines the relationship among these groups.

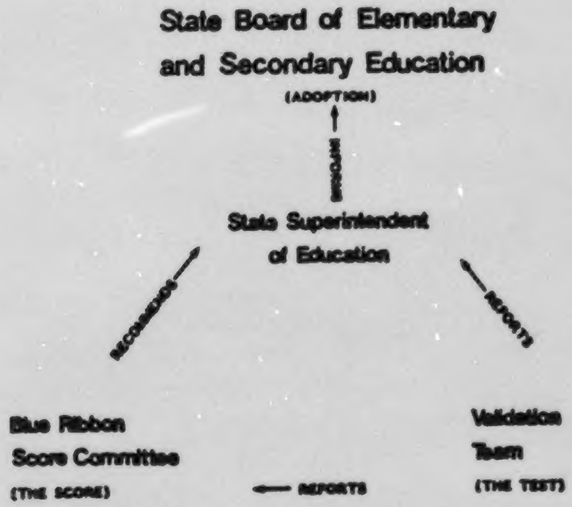
State Board of Elementary and Secondary Education
(ADOPTION)
INFORMS
State Superintendent of Education
RECOMMENDS
REPORTS
Blue Ribbon Score Committee
(THE SCORE)
REPORTS
Validation Team
(THE TEST)

FIGURE 1. VALIDATION DECISION FLOW

State Superintendent of Education

The Superintendent is responsible for selecting a test to be used in teacher certification and establishing the required score on that test. He refers his decision to the State Board of Elementary and Secondary Education. The Superintendent forms his decision after receiving a validation study report and the recommendations of the Blue Ribbon Score Committee about a required test score.

Blue Ribbon Score Committee

This committee was appointed by the Superintendent of Education to recommend a required score on the NTE. The composition of the committee membership met the requirements of R.S. 17:7(6) about those persons whose recommendations the Superintendent is to consider. The Blue Ribbon Score Committee included teachers, principals, a local superintendent of schools, representatives of teachers' organizations, deans of schools of education, members of the governing boards for higher education, and persons who represented the State Board of Elementary and Secondary Education, business, labor, and civic groups. A complete listing of the committee members is given in Appendix I-A.

The Blue Ribbon Score Committee has two functions. The first is to recommend to the Superintendent which modules (all or some) of the Core Battery are sufficiently valid for use in Louisiana. The second function is to recommend a minimal performance level required for teacher certification. In performing the latter function the committee must decide upon both required performance on the Core Battery and the composition of the score or scores (weighted or unweighted, single for the entire NTE or expressed as a multi-component score). The validation study reported here is a major source of information for the Blue Ribbon Score Committee in making its recommendation.

Validation Team

The Validation Team members were nominated by deans of colleges of education from public and private institutions of higher education in Louisiana. They were nominated at the request of the State Department of Education, and these faculty members acted as judges in validating the Core Battery. The selection and composition of the Validation Team is fully described in Chapter II, and its role is discussed later in this chapter and throughout the report. The members are listed in Appendix I-B. The judgments of the Validation Team provided the data for the Content Review and Knowledge Estimation components of this report.

Technical Support

Educational Testing Service conducted the 1978 NTE validation study in a contractual agreement with the Louisiana Department of Education. The Superintendent assigned responsibility for the 1983 study to the Department of Education, Office of Research and Development. Staff from this office was reponsible for designing and conducting the study, analyzing the results, and writing the final report. Other Offices and Bureaus within the Department, particularly the Bureau of Higher Education and Teacher Certification, also served in managing the study and providing consultation. Department of Education staff members involved in the study are listed in Appendix I-C.

Educational Testing Service maintained a consulting role under contract in the 1983 validation. This group provided procedural recommendations and the test items needed for the validation study, attended the Validation Team meetings to ensure security of Educational Testing Service materials, and permitted special administrations of the Common Examinations in Louisiana in November, 1982 and March, 1983.

STUDY DESIGN AND RATIONALE

It should be noted that, while the NTE is required for certification as a beginning teacher in Louisiana, it is not the sole requirement for certification. Therefore, this study is concerned only with the validation of the NTE and does not address any other aspect of certification. This study was based upon the design that had been developed by Educational Testing Service and used in other states as well as in the 1978 Louisiana validation of the NTE. The design and procedures are described fully in A Manual for Doing Content Validity Studies the National Teacher Examinations for Certification Uses (Educational Testing Services, undated). The preface to this manual argues legal and professional justification for the validation procedures.

> The NTE Policy Council's guidelines for using the National Teacher Examinations state:
>
> > In keeping with the sentiment expressed in the decision rendered in the U.S. v. South Carolina (South Carolina 1977) case the Council requires that an NTE user develop a rationale that states the place of the examinations in that certification process. Further, the Council requires that a validity study be conducted to determine the relationship between the teacher training curricula of the state and the outline and specific test questions for each test being considered. In the process of conducting the validity study, if a minimum score is to be established,

> it should be established in relation to some criterion, such as the judgments of experts.
>
> This policy is consistent with guidelines for test use that have been promulgated by other groups and agencies (American Psychological Association, the National Council for Measurement in Education, and the Educational Testing Service, for example) as well as with relevant decisions rendered in the federal courts. (Educational Testing Service, undated, Preface)

Use of the NTE

Professional opinion in the field of education is not unanimous about the use of a test for teacher certification. The disagreements about the philosophy of this practice have been articulated by two major teachers' organizations. The American Federation of Teachers feels that it is approporiate and desirable to test teacher candidates to ensure that they meet minimal standards, and "welcomes fair and valid teacher competency tests" (Shanker and Ward, 1982, p. 8). The National Education Association, however, endorses the rigorous evaluation of aspiring teachers, but feels that testing at the completion of a teacher education program is ill-timed and inappropriate (Hodgkins and McKenna, 1981). These authors argue that proficiency in the basic skills and liberal arts should be assessed before a student has invested four years in a teacher education program. They have two further objections: a) that schools of education and state licensing agencies are too far removed from the acquisition of skills and knowledge in specialty areas to assess these adequately, and b) that the paper and pencil tests that are frequently used are generally too narrow and inaccurate. In the judgment of those responsible for this study the last set of arguments does not apply to evaluating the validity of the NTE for teacher certification in Louisiana. First, the study does not address the point in a student's educational career at which the NTE is administered. Second, the NTE is considered a measure of knowledge possessed by a teacher candidate, not a measure of teaching performance. Hence, the college faculty members responsible for preparing teacher candidates are appropriate judges of the NTE's validity as a measure of the teacher education curricula. The Louisiana legislature has already decided that certain areas of teacher candidate knowledge shall be assessed.

Beginning as early as 1976 four states required passing the NTE as a condition for teacher certification (Vlaanderen, 1982). An additional 23 states have used the NTE for certification in specific areas or for validating credits from an unaccredited institution. Eight states currently use the NTE in teacher certification: Alabama, Arkansas, Louisiana, Mississippi, South Carolina, Tennessee, Virginia, and West Virginia.

In summary, common practice and legal decisions to date permit the use of the NTE for teacher certification when appropriate validation studies have been employed. The methodology used for the validation study reported here is felt to be appropriate because it is, with minor additions, the one developed by the publisher of the NTE, and the test has withstood legal challenge when this validation procedure is employed. The present study has added two components that, in the judgment of the researchers, contributed important information about test content and improved the quality of the validation; these are discussed briefly in the design section and in the following chapters describing the results.

Design

The study employed the jury judgment approach used in the 1978 NTE validation study conducted by Educational Testing Service. In this method, panels of faculty members were drawn from Louisiana institutions of higher education to a) validate the content of the tests in the Core Battery and b) estimate the score that could be expected from a minimally knowledgeable teacher candidate. Educational Testing Service had selected a jury judgment approach over other alternatives in 1978 because none of these could provide equally extensive and reliable data within reasonable time and staffing constraints (Educational Testing Service, 1978, pp. 42-43). The rejected alternatives--a detailed content analysis of written documents or observation of NTE scores among a group of teacher candidates independently judged to possess a minimally acceptable amount of knowledge--appeared equally impracticable for the 1983 study.

Figure 2 outlines the design for the study. Faculty members who met qualifications concerning teaching areas and years of experience in Louisiana colleges and universities were nominated by deans of colleges of education. Panelists were then selected from this pool to form a group that was representative of Louisiana institutions in three areas:

- the number of teacher education majors graduating from the universities;
- the proportion of public and private colleges and universities;

- the proportion of institutions with predominantly white and predominantly black student enrollments.

The selection and composition of Validation Team panels are fully discussed in Chapter II.

Validation Team members were assigned to serve on either a Content Review Panel or a Knowledge Estimation Panel, and they judged tests that were appropriate to the respective subject areas from which they were nominated.

Each Content Review panelist made two major judgments about the Core Battery test to which he or she was assigned.

- Is the emphasis given to topics within the test the same as the emphasis given to those topics in the Louisiana teacher education curricula with which the panelist is familiar?

- Would 90 percent or more of the graduates from the teacher education programs in Louisiana with which the panelist is familiar have had the opportunity to learn the content included in each test item?

The panelists' judgments about the congruence between test topics and curriculum content determined how well the tests matched teacher education programs. Their judgments about the appropriateness of items (whether students would have had the opportunity to learn the required content) was combined with this to evaluate the overall content validity of each test for Louisiana. Only those items judged to be appropriate by the Content Review panelists were included in the analysis of the Knowledge Estimation results. In other words, an item judged to be inappropriate could not contribute to the recommended score.

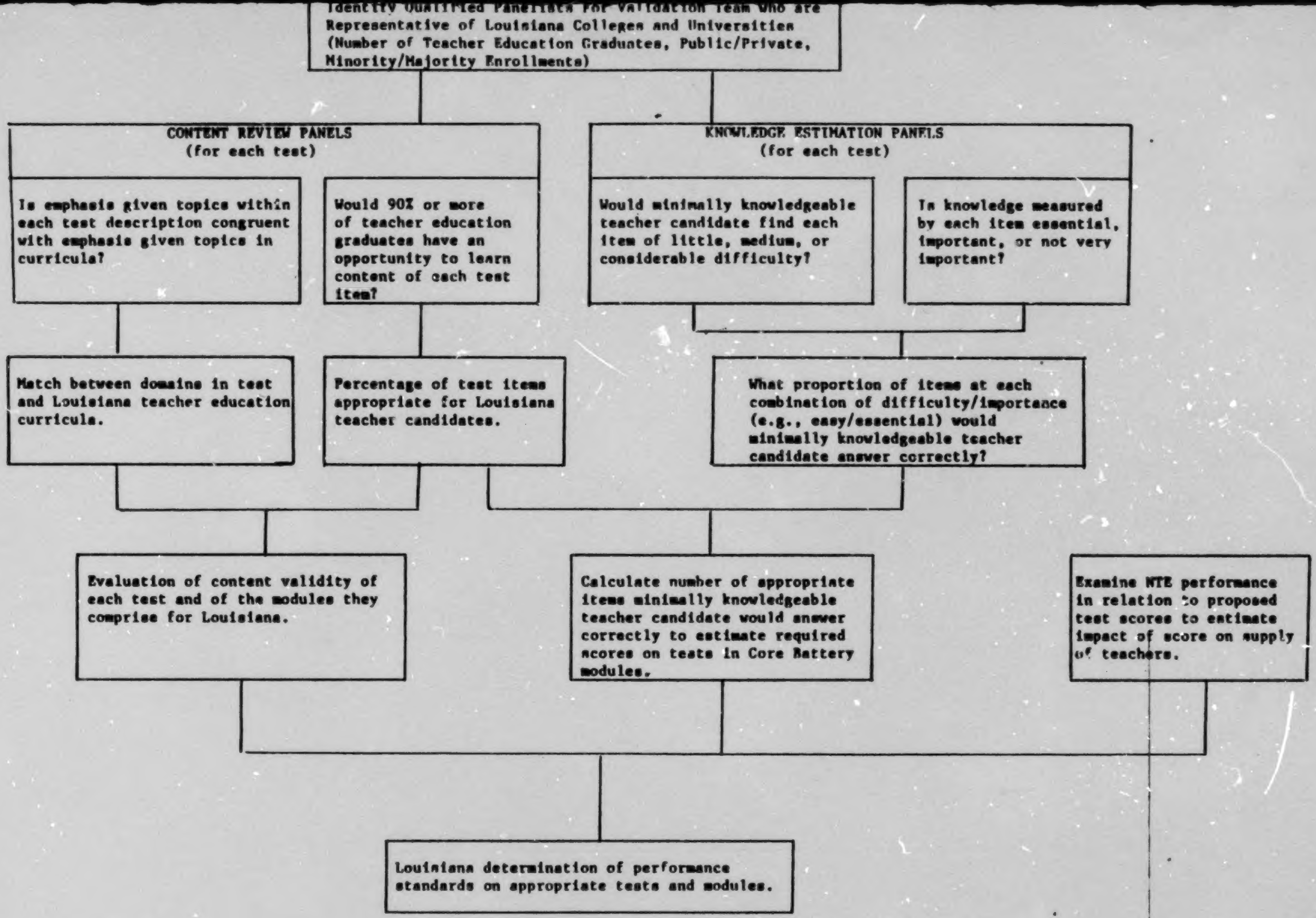
Identify Qualified Panelists for Validation Team Who are Representative of Louisiana Colleges and Universities (Number of Teacher Education Graduates, Public/Private, Minority/Majority Enrollments)
CONTENT REVIEW PANELS (for each test)
KNOWLEDGE ESTIMATION PANELS (for each test)
Is emphasis given topics within each test description congruent with emphasis given topics in curricula?
Would 90% or more of teacher education graduates have an opportunity to learn content of each test item?
Would minimally knowledgeable teacher candidate find each item of little, medium, or considerable difficulty?
Is knowledge measured by each item essential, important, or not very important?
Match between domains in test and Louisiana teacher education curricula.
Percentage of test items appropriate for Louisiana teacher candidates.
What proportion of items at each combination of difficulty/importance (e.g., easy/essential) would minimally knowledgeable teacher candidate answer correctly?
Evaluation of content validity of each test and of the modules they comprise for Louisiana.
Calculate number of appropriate items minimally knowledgeable teacher candidate would answer correctly to estimate required scores on tests in Core Battery modules.
Examine NTE performance in relation to proposed test scores to estimate impact of score on supply of teachers.
Louisiana determination of performance standards on appropriate tests and modules.

FIGURE 2. STUDY DESIGN

BEST COPY AVAILABLE

The Knowledge Estimation panelists were asked to determine what they considered to be the minimally knowledgeable teacher candidate: a person possessing the minimal amount of academic knowledge needed to a) complete the college program required for certification in Louisiana and b) teach effectively. Using this minimally knowledgeable teacher candidate as a frame of reference, each panelist then answered three questions about the test to which he or she had been assigned:

- Would the minimally knowledgeable teacher candidate find each item easy, moderately difficult, or hard?
- Is the knowledge measured in each item essential, important, or not very important?
- What proportion of items would the minimally knowledgeable teacher candidate answer correctly in each of the nine categories of difficulty and importance (easy/essential, easy/important, easy/not very important, etc)?

The ratings of importance were an addition to the 1978 study design. In that earlier validation, judges had been asked to rate only item difficulty. Including ratings of item importance provided an additional dimension to the measurement of validity. The rationale for this change in the study design is discussed in Chapter IV.

The Knowledge Estimation panelists' judgments of the number of items in each of the nine difficulty and importance categories were combined with their judgments of the percentage of items in each category a minimally knowledgeable teacher candidate would answer correctly. These data were further combined with the judgments about item appropriateness furnished by the Content Review Panels to calculate the number of appropriate items the minimally knowledgeable teacher candidate would answer correctly on the Core Battery Tests.

The last element of the validation study, involved a second addition to the 1978 validation procedures. This addition was the opportunity for teacher candidates who would be eligible for certification by September 15, 1983, and who took the Core Battery in November, 1982, or March, 1983, to also take the Common Examinations at no extra cost. The Common Examinations were no longer in use by Educational Testing Service after the introduction of the Core Battery. Allowing Louisiana teacher candidates to take both examinations was a concession on the part of the testing company. The purpose of the double testing option was to avoid penalizing any teacher candidate who wished to apply for certification before the new NTE score was established in July, 1983.

As the last step in Figure 2 shows, three sets of information were provided to assist in recommending and determining a required score on the NTE. These were:

- Evaluation of the content validity of Core Battery Tests, judged by the Content Review Panels;
- A score that could be expected of the minimally knowledgeable teacher candidate, formed from the judgments of the Content Review and Knowledge Estimation Panels; and
- Performance information to be used in evaluating the effect of different qualifying scores on Louisiana teacher candidates attempting the Core Battery.

REFERENCES CITED

Educational Testing Service. A Manual for Doing Content Validity Studies the National Teacher Examinations for Certification Uses. Princeton, New Jersey: Educational Testing Service, undated.

Educational Testing Service. Report on a Study of the National Teacher Examinations by the State of Louisiana. Princeton, New Jersey: Educational Testing Service, 1978.

Hodgkins, R., and McKenna, B. Testing and teacher certification an explosive combination. Educational Measurement, 1982, 1(2), 10 ff.

Shanker, A., and Ward, J.G. Teacher competency and testing: a natural affinity. Educational Measurement, 1982, 1(2), 6 ff.

Vlaanderen, R.B. Teacher competency testing: status report. Educational Measurement, 1982, 1(2), 17 ff.

West's Louisiana Revised Statutes. 17:7(6).

CHAPTER II

PANEL SELECTION

PANEL SELECTION PROCESS

The Validation Team members who formed the Content Review and Knowledge Estimation Panels were selected to represent equitably the Louisiana colleges and universities that have teacher education programs. There were three objectives considered in the composition of these panels: a) designating members with the required expert qualifications, b) representing teacher training programs fairly by size and public/private status of institutions, and c) representing colleges and universities with significant black student enrollments fairly.

The first objective required that the college faculty members who were nominated as panelists have the desired length and kind of teaching experience. Nominations were to be limited to faculty who had taught for two or more years in a Louisiana postsecondary institution, and who had taught a course within the last five years. The colleges and universities were also requested to nominate faculty from the professional areas or fields that paralleled the content of Core Battery tests. The fields for the tests within each module of the Core Battery are outlined below.

- Professional Knowledge: there was one test within this module. Faculty members were to be nominated from the academic fields of educational psychology, instructional methodology, measurement and evaluation, and the sociology, organization, and administration of schools.

- General Knowledge: there were four tests in this module. The areas for faculty nominations to each test were:
  Science Test: biology, physical science, chemistry, astronomy, geology, meteorology, and general science.

Social Studies Test: history, political science, economics, sociology, anthropology, and geography.
Literature/Fine Arts Test: English, music, and art.
Mathematics Test: mathematics and mathematics education.

- Communication Skills: there were three tests in this module. The areas for faculty nominations to each test were:
  Reading Test: English.
  Writing Test: English.
  Listening Test: speech and communications.

The second objective required that each institution provide a proportional number of panelists for the study based upon the number of teacher education graduates in 1980-1981. Institutions were asked to nominate a specific number of panelists determined by the relative size of their teacher education programs. For example, if a university had graduated 10 percent of the teacher education majors in 1980-1981, it was asked to nominate 10 percent of the panelists needed for the validation study.

The third objective was to ensure fair representation of universities and colleges with significant black enrollments. The State Department of Education collected information about the proportion of ethnic minorities within the student body of each institution having a teacher education program. Each institution was then categorized as predominantly (51% or more) black or white. The number of panelists requested from each was then determined in a manner to ensure that the proportion of nominations from predominantly black or white institutions matched the proportion of students from these institutions. For example, if 25 percent of the State's college and university students were enrolled in predominantly black institutions, then 25 percent of the panelists were to be nominated from these schools. It should be noted that this was based upon the composition of the student body, not the ethnic background of the faculty members nominated. Predominantly white institutions were free to nominate nonwhite panelists, and vice versa.

The President of each of the 22 institutions with teacher training programs was asked to designate a Campus Coordinator to manage the panel nominations. Each Campus Coordinator was given a form listing the areas or fields from which nominees were requested and the number of faculty members required for each area. He or she was then asked to send the Bureau of Research, Office of Research and Development, a list of panel nominees with background information on each. The Assistant Director of this Bureau was responsible for chairing the task force that conducted this study. The information included the number of years the nominee had taught in Louisiana, the major and minor fields of the person's degree, membership in professional organizations, and a list of courses taught in the last five years. It could not be expected that all faculty members would have taught all of the topics within the Core Battery tests (Science, for example, included eight disciplines). Thus, the faculty background data were used to select panelists who would represent a broad range of the required expertise.

PANEL NOMINATIONS

Table 1 shows the number of faculty members requested and nominated from each of the institutions. Three hundred seven nominations were requested and 302 were received. Four universities nominated more

TABLE 1

NUMBER OF FACULTY MEMBERS REQUESTED (REQ) AND NOMINATED (NOM); SHOWN BY INSTITUTION AND AREA

	Professional Knowledge		Math		Science		Social Studies		Art		Music		English		Speech		TOTAL	
	REQ	NOM	REQ	NOM	REQ	NOM	REQ	NOM	REQ	NOM	REQ	NOM	REQ	NOM	REQ	NOM	REQ	NOM
aptist Christian	1	1											1	1			2	2
entenary	1	1	0	1									1	1	1	0	3	3
illard	1	1	1	0			0	2					1	0			3	3
rambling	4	4	3	2	3	3	3	4	1	1	1	1	6	6	3	3	24	24
ouisiana College	1	1	1	1	1	1	1	1	1	0	1	1	1	0			7	5
SU-Baton Rouge	4	4	4	4	3	3	3	3	2	2	1	1	10	10	4	4	31	31
SU-Shreveport	1	1	1	1	1	1	1	2					1	1	1	1	6	7
ouisiana Tech	3	3	1	1	1	1	1	1					4	4	1	1	11	11
oyola	1	1	1	2	1	0							1	1			4	4
cNeese	4	4	3	3	3	3	3	3	1	1	1	1	5	5	3	3	23	23
icholls	4	4	3	3	3	3	3	4	1	0	1	1	6	6			21	21
ortheast Louisiana	4	4	3	3	3	3	3	4			1	1	5	5	2	2	21	22
orthwestern State	3	3	2	1	2	2	2	2	1	1	1	1	2	2	2	1	15	13
Our Lady Holy Cross	1	1					1	1					1	1	1	1	4	4
St. Mary's Dominican	1	1	1	1			1	1					1	1			4	4
Southeastern University	5	5	3	4	3	4	3	3	1	1	1	2	6	7	3	3	25	29
Southern-Baton Rouge	5	6	4	4	4	4	5	5	1	1	1	1	11	10	4	4	35	35
Southern-New Orleans	1	1	1	1	1	1			1	1	1	0	3	3			8	7
Tulane	1	1			1	1							1	1			3	3
University of New Orleans	4	5	2	2	3	3	3	3	1	0	1	1	8	6	3	0	25	20
University Southwest LA	4	4	2	2	3	3	3	4	1	1	1	1	10	10	5	6	29	31
Xavier	1	0											1	0	1	0	3	0
Total Nominees	55	56	36	36	36	36	36	43	12	9	12	12	86	81	34	29	307	302

faculty members than had been requested, and four submitted fewer nominees than requested. Xavier University did not provide any nominees and did not participate in the validation study.

Within the subject areas of the tests, 90 percent or more of the number of nominations requested were received in all but two areas. These were art, in which 75 percent of the requested nominations were received, and speech/communications, in which 85 percent of the requested nominations were received.

The number of nominations requested and received was actually greater than the number of panelists required. The study design called for a total of 36 panelists in Professional Knowledge to form Content Review and Knowledge Estimation Panels composed of 16 members and two alternates each. All of the other test areas were to have 24 panelists: 10 members and two alternates for both the Content Review and Knowledge Estimation Panels. All panelists, including the alternates, were to participate in the validation study. The excess nominations were requested to allow for a proper balance of expertise among panelists, the possibility of faculty members withdrawing from a panel, and nonparticipation because of unforeseeable schedule conflicts. Table 2 shows the size of the panels drawn for each test area and the number of panelists selected from each institution. A total of 205 panelists was selected from the 302 nominees. Thirty-seven panelists were selected for Professional Knowledge, and 24 were selected for each of the other test areas. The additional panelist in Professional Knowledge was nominated by a predominantly black private university and was selected to

TABLE 2

ALLOCATION OF FACULTY MEMBERS SELECTED, SHOWN BY INSTITUTION AND TEST

	Professional Knowledge	Math	Science	Social Studies	Literature Fine/Arts	Listening	Reading	Writing	TOTAL
Baptist Christian	1								1
ntenary	1	1							2
Dillard	1			2					3
rambling	4	2	1	1	3	3	2	1	17
Louisiana College	1	1	1	1	1				5
LSU-Baton Rouge	2	2	2	2	4	3	3	2	20
SU-Shreveport			1	1		1	1		4
Louisiana Tech	1	1	1			1	1	2	7
yola	1	1			1				3
McNeese	2	2	1	2	2	3	2	2	16
icholls	2	2	3	3			3	2	15
Northeast Louisiana	2	2	1	3	2	2	1	1	14
rthwestern State	2	1	1	1	1	1		2	9
Our Lady Holy Cross	1						1		2
. Mary's Dominican	1	1						1	3
utheastern University	3	2	3	2	2	2	1	2	17
uthern-Baton Rouge	4	2	2	2	3	3	3	3	22
uthern-New Orleans	1	1	1		2		1	1	7
Tulane	1		1						2
iversity of New Orleans	4	2	3	2	2		2	2	17
University Southwest LA	2	1	2	2	1	5	3	3	19
tal Nominees	37	24	24	24	24	24	24	24	205

compensate somewhat for Xavier University's nonparticipation. Within the constraints of the nominees' expertise, every attempt was made to distribute faculty from different institutions across test areas.

Representativeness of Panels

Table 3 presents the proportion of 1980-1981 Louisiana teacher education graduates and the proportion of faculty from each university among the panelists who were nominated, selected, and who actually participated in the study. There were only two cases in which the percent of faculty selected for the panels differed by more than half a percentage point (.005) from the percentage of the State's teacher education graduates coming from that institution. These were Southern University--Baton Rouge, which produced 11.5 percent of the teacher education graduates and from which 10.7 percent of the panelists were selected, and Southern University--New Orleans, which produced 2.5 percent of the State's teacher education graduates and from which 3.4 percent of the panelists were selected. The percentages of panelists participating by institution differed somewhat from the percentages of panelists selected because some selected faculty members were not able to attend the validation study meetings. However, the discrepancies here were also small. The greatest differences were for Grambling State University, which provided 8.3 percent of the selected panelists and 6.8 percent of the participating panelists, and Nicholls State University, with 7.3 percent of the panelist selections and 7.9 percent of the participating panelists.

Table 4 describes the panel selections in terms of the predominant racial background of the student bodies in the participating institu-

TABLE 3

PERCENTAGE OF GRADUATES AND FACULTY NOMINATED, SELECTED, AND PARTICIPATING, BY INSTITUTION

Institutions	1980-1981 Graduates	Nominated	Faculty Selected	Participating[a]
Baptist Christian	0.6	0.7	0.5	0.5
Centenary	0.8	1.0	1.0	1.0
Dillard	1.1	1.0	1.5	1.6
Grambling	8.0	7.9	8.3	6.8
Louisiana College	2.4	1.7	2.4	2.6
LSU-Baton Rouge	10.0	10.3	9.8	9.9
LSU-Shreveport	1.9	2.3	2.0	2.1
Louisiana Tech	3.6	3.6	3.4	3.1
Loyola	1.3	1.3	1.5	0.5
McNeese	7.6	7.6	7.8	8.4
Nicholls	7.0	7.0	7.3	7.9
Northeast Louisiana	6.8	7.3	6.8	6.8
Northwestern State	4.9	4.3	4.4	4.7
Our Lady Holy Cross	1.2	1.3	1.0	1.0
St. Mary's Dominican	1.3	1.3	1.5	1.6
Southeastern University	8.2	9.6	8.3	8.4
Southern-Baton Rouge	11.5	11.6	10.7	11.0
Southern-New Orleans	2.5	2.3	3.4	3.1
Tulane	0.8	1.0	1.0	1.0
University of New Orleans	8.1	6.6	8.3	8.4
University Southwest LA	9.4	10.3	9.3	9.4
Xavier	0.9	0.0	0.0	0.0
TOTAL(N)[b]	99.9	100.0	100.2	99.8

---

[a] Includes nine qualified substitutes who replaced originally selected panelists.

[b] Percents may not total 100 due to rounding.

tions. It also shows the proportions of graduates and panelists from public and private institutions. Overall, the schools with a predominantly black student enrollment graduated 24.0 percent of the State's teacher education majors; 23.4 percent of the selected panelists came from these institutions, as did 22.5 percent of the panelists who actually participated in the study. State-supported institutions produced 89.5 percent of the 1980-1981 teacher education graduates. Some 90.1 percent of the selected panelists and 89.9 percent of the participating panelists were from these schools.

The sex and years of experience of the faculty members who were nominated and selected, and who participated in the study, are presented in Table 5. The proportion of women faculty members participating (31.9%) did not differ greatly from the proportion of women among the panelists nominated (34.8%). Three faculty members were nominated who did not have the required two years of experience. These persons were not selected as panelists. The majority of the faculty nominated (74.1%) and participating (72.3%) had 10 or more years of teaching experience.

In summary, the panelists who were nominated and selected, and who participated in the validation study, were representative of teacher education programs in Louisiana. The selection process produced no apparent bias in terms of the size of the teacher education program, or in the ethnic identification or public/private status of the universities represented. The selected faculty also matched the total group of nominees in sex distribution and years of experience.

TABLE 4

FACULTY MEMBERS NOMINATED, SELECTED, AND PARTICIPATING FROM INSTITUTIONS CLASSIFIED BY TYPE OF SUPPORT AND PREDOMINANT RACIAL COMPOSITION OF STUDENT BODY: PERCENTAGES

Institutional Classification	1980-1981 Graduates (Percent)	Faculty Members			
		Requested (Percent)	Nominated (Percent)	Selected (Percent)	Participating (Percent)
State Supported					
Predominantly Black	22.0	21.8	21.9	22.4	20.9
Predominantly White	67.5	67.3	68.9	67.3	69.0
Private					
Predominantly Black	2.0	2.0	1.0	1.5	1.6
Predominantly White	8.4	8.8	8.3	8.8	8.4
All State Supported	89.5	89.1	90.7	90.1	89.9
All Private	10.4	10.8	9.3	9.8	10.0
All Predominantly Black	24.0	23.8	22.8	23.4	22.5
All Predominantly White	75.9	76.1	77.2	76.5	77.4
Base Number	4285	307	302	205	191

TABLE 5

FACULTY MEMBERS NOMINATED, SELECTED, AND PARTICIPATING BY SEX AND YEARS OF EXPERIENCE TEACHING IN LOUISIANA: NUMBERS AND PERCENTAGES

Characteristic	Faculty					
	Nominated		Selected		Participating	
	n	%	n	%	n	%
Sex						
Female	105	34.8	66	32.2	61	31.9
Male	197	65.2	139	67.8	130	68.1
	n	%	n	%	n	%
Years of Experience						
Less than 2	3	1.0	0	0	0	0
2 to 5	35	11.6	28	13.7	28	14.7
6 to 9	40	13.2	25	12.2	25	13.1
10 or more	212	70.2	152	74.1	138	72.3
Unknown	12	4.0	0	0	0	0
Base Number	302		205		191	

PANEL ASSIGNMENTS

The selected Validation Team members were randomly assigned to either a Content Review or a Knowledge Estimation Panel that was appropriate to their area of expertise. Panelists were notified of their assignments by their Campus Coordinators and were encouraged to familiarize themselves with their institution's total offerings in the area of the test they were to evaluate. To help in this preparation, the panelists were also mailed rating task directions and instructions and test content descriptions a month prior to the panel assemblies.

Content Review Panel Assignments and Attendance

Table 6 shows the number of Content Review panelists selected to evaluate each test and the college or university from which these panelists were nominated. The table also reports the number who actually attended the October 13, 1982 panel assemblies.

The Writing Test panelists evaluated the objective subtest on October 13, 1982, and essay subtest on February 17, 1983. The latter session was necessary because the panelists worked with samples of actual essays from the November, 1982, Louisiana administration of the Core Battery. The second form of the Listening Test was also not available in October, and the Listening panelists were called back on February 10 and 11, 1983, to evaluate that form.

Content Review panelists were selected from 20 of the 21 universities participating in the validation study, and were in attendance from all but one of these institutions. A total of 102 panelists was selected, and 94 (92%) attended. The smallest number of panelists

TABLE 6

NUMBER OF FACULTY MEMBERS SELECTED AND ATTENDING CONTENT REVIEW PANEL ASSEMBLY, SHOWN BY TEST JUDGED AND INSTITUTION

	PROFESSIONAL KNOWLEDGE		GENERAL KNOWLEDGE								COMMUNICATION SKILLS						
			Math		Science		Social Studies		Literature Fine Arts		Reading		Listening		Writing[a]		
																Objective	Essay
	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Attend
Baptist Christian	1	1															
Centenary			1	1													
Dillard																	
Grambling	2	2	1	1			1	1	2	1	1	1	1	1			
Louisiana College			1	1													
LSU-Baton Rouge	1	1	1	1	1	1	1	1	2	2	1	1	1	1[b]	1	1	0
LSU-Shreveport					1	1	1	1									
Louisiana Tech					1	0					1	1			1	1	1
Loyola	1	0							1	0							
McNeese	1	1	1	1			1	1	1	1	1	1	2	1	1	1	1
Nicholls	1	1	1	1	2	2	1	1			1	1			1	1	1
Northeast Louisiana	1	1	1	1			1	1	1	1	1	1	1	1			
Northwestern State	1	1							1	1			1	1	1	1	1
Our Lady Holy Cross	1	1															
St. Mary's Dominican			1	1											1	1	1
Southeastern University	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1
Southern-Baton Rouge	2	2	1	1	1	1	1	0	2	2	1	1	2	2	2	2	2
Southern-New Orleans	1	1	1	0	1	1					1	1			1	1	1
Tulane	1	1			1	1											
University of New Orleans	2	2	1	1	2	1	1	1	1	1	1	1			1	1	1
University Southwest LA	1	1			1	1	2	2			2	2	3	3	1	1	1
Total	18	17	12	11	12	10	12	11	12	10	12	12	12	11	12	12	11

TOTAL SELECTED: 102

TOTAL ATTENDING: 94

---

[a] The same panelists evaluated the objective and essay subtests in the Writing Test

[b] Two forms of Listening Test evaluated in separate sittings; one panelist did not return for review of second form.

attending for any test was 10 (Science and Literature/Fine Arts Tests). Attendance was equal to or greater than the number called for in the design for each of the other tests and compared favorably with that in the 1978 validation study.

Knowledge Estimation Panel Assignments and Attendance

The numbers of panelists selected for and attending the Knowledge Estimation Panel assembly are shown in Table 7 by institutional affiliation. The Knowledge Estimation Panel met on October 14 and 15, 1982. The first day of the meeting consisted of training sessions to familiarize the participants with the validation tasks they would perform on the following day. Panelists evaluating the essay subtest of the Writing Test and the second form of the Listening Test were reconvened in February, 1983, along with Content Review Panelists for these tests.

Twenty of the universities taking part in the validation study contributed faculty members to the Knowledge Estimation Panel selection, and panelists attended from each of these schools. A total of 103 panelists was selected. Ninety-seven of them (94%) attended and 93 (90%) provided data that were used in the study. Four panelists were disqualified because they inadvertantly failed to meet the procedural requirements of the study. These panelists do not appear in the Knowledge Estimation results. The smallest attendance was for the essay subtest of the Writing Test (seven panelists attended) and the Listening Test (nine reviewed the first form, seven reviewed the second). These numbers were below those called for in the study design but were considered sufficiently large to allow validation analyses of these tests with the

TABLE 7

NUMBER OF FACULTY MEMBERS SELECTED AND ATTENDING KNOWLEDGE ESTIMATION PANEL ASSEMBLY, SHOWN BY TEST JUDGED AND INSTITUTION

	PROFESSIONAL KNOWLEDGE		GENERAL KNOWLEDGE								COMMUNICATION SKILLS						
			Math		Science		Social Studies		Literature Fine Arts		Reading		Listening		Writing[a]		
																Objective	Essay
	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Select	Attend	Attend
Baptist Christian																	
Centenary	1	1															
Dillard	1	1					2	2									
Grambling	2	2	1	0	1	1			1	1	1	1	2	0	1	1	0
Louisiana College	1	1			1	1	1	1	1	1							
LSU-Baton Rouge	1	1	1	1	1	1	1	1	2	2	2	2	2	1	1	1	0
LSU-Shreveport											1	1	1	1[b]			
Louisiana Tech	1	1	1	1									1	1	1	1	0
Loyola			1	1													
McNeese	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Nicholls	1	1	1	1	1	1	2	2			2	2			1	1	1
Northeast Louisiana	1	0	1	1	1	1	2	2	1	1			1	1	1	1	0
Northwestern State	1	1	1	1	1	1	1	1							1	1	1
Our Lady Holy Cross											1	1					
St. Mary's Dominican	1	1															
Southeastern University	2	2	1	1	2	1			1	1			1	1	1	1	1
Southern-Baton Rouge	2	2	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1
Southern-New Orleans									2	2							
Tulane																	
University of New Orleans	2	2	1	1	1	1	1	1	1	1	1	1			1	1	0
University Southwest LA	1	0	1	1	1	1			1	1	1	1	2	2[b]	2	2	2
Total	19	18	12	11	12	11	12	12	12	12	12	12	12	9	12	12	7

TOTAL SELECTED: 103

TOTAL ATTENDING: 97

---

[a] The same panelists evaluated the objective and essay subtests in the Writing Test.

[b] Two forms of Listening Test evaluated in separate sittings; two panelists did not return for review of the second form.

Knowledge Estimation Panel. In the case of the Writing essay subtest, 11 Content Review panelists also performed the Knowledge Estimation rating tasks. This was due to the nature of the essay subtest, which differed in form from the other Core Battery Tests, and for which formal Content Review ratings would not have been meaningful. The greater of number of ratings thus increased the stability of the data. Attendance was above the number desired for all other tests.

HALF-PANELS

After the October rating sessions, each attending panelist was assigned to a half-panel. These half-panels were formed to measure the consistency of responses among judges on the Content Review and Knowledge Estimation tasks. Half-panels never met as a group or performed any duties separately.

The two half-panels for each test were constituted to be as similar as possible to one another with regard to average tenure, representation of predominantly black universities, and the departments in which the judges taught. This was done to form half-panels that would allow a comparison of ratings by the two groups that would not be expected to differ because of these factors. In measuring consistency for each test, data from the two half-panels were treated as if the groups had given independent ratings on the Content Review and Knowledge Estimation tasks. The analyses of the half-panel data in Chapters III and IV are measures of intragroup (half-panel) reliability, i.e., a measure of the consistency of the ratings. Rating tasks on which the half-panel ratings do not agree consistently may indicate a need for caution in interpreting the results of that analysis.

Selection and Composition of Half-Panels

Judges were allocated to half-panels on the basis of three criteria: a) the number of years each had taught in Louisiana, b) the racial composition of the student body at the university in which the judge taught, and c) the department in which the judge taught. The half-panels that resulted from this selection process were deliberately similar to each other. Table 8 compares the half-panels for each test on size, average tenure of members, and the proportion of panelists from predominantly black institutions.

In can be noted that the average tenure of the two half-panels for any test did not differ by more than 2.8 years. The mean difference in tenure between the half-panels for a test was 0.9 years. Similarly, the half-panels were constructed so that each included judges from predominantly black institutions when possible. Only three of the sixteen half-panels did not include one or more judges from a predominantly black institution. These three exceptions were instances in which half-panels had been constructed from Panels that contained only one judge from a predominantly black institution, and that judge was arbitrarily placed in the other half-panel of the pair.

Finally, the department in which each judge taught was considered when half-panels were selected. This was especially important for the Social Science and Science Panels, in which a variety of disciplines was represented. The assignments were based upon distributing judges from the same disciplines, indicated by the university departments in which they taught, to different half-panels within each test.

TABLE 8
NUMBER AND CHARACTERISTICS OF JUDGES ON CONTENT REVIEW AND KNOWLEDGE ESTIMATION HALF-PANELS, BY TEST

	CONTENT REVIEW						KNOWLEDGE ESTIMATION					
		Panel 1			Panel 2			Panel 1			Panel 2	
	Number	Average Years of Tenure	Number For Black Institution	Number	Average Years of Tenure	Number for Black Institution	Number	Average Years of Tenure	Number for Black Institution	Number	Average Years of Tenure	Number for Black Institution
Professional Education	9	18.0	2	8	17.1	2	8	16.0	2	8	16.0	2
Mathematics	6	14.8	1	5	13.8	1	5	14.8	0	5	15.0	1
Science	5	15.8	1	5	17.0	1	5	15.4	1	6	14.8	1
Social Studies	5	11.2	0	6	12.0	1	6	13.8	1	6	12.0	2
Literature/Fine Arts	5	10.4	2	5	11.8	1	5	16.0	2	5	13.2	2
Reading	6	15.7	1	6	15.0	2	6	14.8	2	6	15.0	1
Listening[a] (Form A)	6	16.5	2	5	14.8	1	5	16.0	1	4	16.3	0
(Form B)	5	16.6	2	5	17.4	1	4	18.5	1	3	17.3	0
Writing (Objective)	6	13.2	2	6	12.8	1	6	16.0	1	5	15.6	1
Writing[b] (Essay)							9	13.4	2	9	13.2	2

[a] Listening Test Knowledge Estimation includes separate half-panels for forms A (upper line) and Form B (lower line) of that test. This procedure accommodates differences in attendance at rating sessions for the two forms.

[b] There was no Content Review Panel for the essay subtest of the Writing Test.

It has been noted that the Knowledge Estimation Panelists for the Listening Test evaluated Form A in October, 1982, and Form B in February, 1983. Only seven of the original nine panelists were able to return for the second rating session. Because of this, it was necessary to construct two separate sets of Knowledge Estimation half-panels, one for each Form of the Listening Test. Those panelists who attended both rating sessions are included in both sets of half-panels. Those who did not attend the second session are excluded from the half-panels for Form B. This procedure was followed to prevent an imbalance in half-panel assignments.

The essay subtest of the Writing Test was not included in the Content Review since the subtest did not include discrete items. Hence half-panels were constructed only for the Knowledge Estimation tasks.

SUMMARY

The information presented in this chapter demonstrates that the Knowledge Estimation and Content Review panelists who evaluated the NTE Core Battery provided a comprehensive representation of teacher education programs in Louisiana. All but one of the State's universities with such programs were proportionally represented among the panelists. These faculty members, as a group, reflected the balance of public or private, and predominantly black or white, universities contained within Louisiana. They also represented a wide range of disciplines through the departments in which they taught, and possessed a broad range of teacher education experience. The panelists, in summary, were highly qualified and equitably selected judges of the validity of the tests they reviewed.

CHAPTER III

CONTENT REVIEW RESULTS

INTRODUCTION

The Content Review phase of the validation study involved four types of judgments about the Core Battery tests: a) the content appropriateness of the test items for students who have completed teacher education programs in Louisiana, b) comprehensiveness, i.e., the extent to which the tests included major content topics, c) the match between the emphasis given content topics in the tests and in teacher education programs, and d) the overall similarity between the content of each Core Battery test and that of Louisiana teacher education programs.

The composition of the Content Review Panel is discussed in Chapter II. The panelists met on October 13, 1982, at which time their task descriptions and instructions were reviewed in detail, and the item and test validation ratings were carried out. An overview of the tasks carried out by the Content Review panelists is included in Appendix III-A. Of the 102 Content Review panelists selected, 94 attended the October panel assemblies. The Listening Test panelists reconvened in February, 1983, to evaluate the second form of that test, and the Writing Test panelists were also called back in February to serve on the Knowledge Estimation Panel for the essay subtest.

Content Appropriateness of Items

Judges used the Question Review Form (Appendix III-B) to evaluate the content appropriateness of each item in the test to which they had been assigned. An item was to be rated as content appropriate if the panelist judged that 90 percent or more of the students in the Louisiana teacher education programs with which he or she was familiar would have had the opportunity to learn the knowledge required to answer the question. This learning could have occurred during the student's teacher education course of study or as part of coursework prerequisite to the teacher education program. Judges were instructed to rate each item as either content appropriate or inappropriate unless they had no basis whatsoever for making the judgment. In the latter case they could assign a "Do Not Know" rating.

The test items were first analyzed individually. Some judges selected the "Do Not Know" option, so not all questions were classified by all judges. The "Do Not Know" responses were treated as neither appropriate nor inappropriate, and were thus not included in the data analysis. The exclusion of the "Do Not Know" responses was based on the assumption that, if these judges had had a basis on which to rate the items, their ratings would have been distributed in the same way as those of the other judges who rated the items as appropriate or inappropriate. This procedure had been followed by Educational Testing Service in the 1978 validation study. It was also designed to exclude any items that were rated (as either appropriate or inappropriate) by fewer than three judges on a panel or half-panel. None of the items from any test required exclusion on this basis; all were rated by three or more judges on both half-panels.

Each item was then classified as either appropriate or inappropriate in the judgment of the total panel evaluating the test in which it appeared. An item was classified as content appropriate if 51 percent or more of the judges gave it this rating. This procedure was identical to that used in the 1978 validation study. In actual practice the criterion was somewhat higher than 51 percent. In the largest panel (Professional Knowledge Test) the majority was nine of 17 panelists, or 53 percent. The majority was six of 10 in the smallest panels, or 60 percent.

Table 9 displays the number and percent of items rated as content appropriate for both forms of each Core Battery test. The number and percentage of items given for the Writing Test exclude the essay subtest. That subtest consisted of a topic on which the examinee was instructed to write an essay and did not include discrete items that could be evaluated for their content appropriateness.

Across tests, the percent of items judged content appropriate ranged from 100 percent (Reading and Social Studies Tests) to 84 percent (Literature/Fine Arts Test). It should be noted that only those items rated as content appropriate by the majority (51% or more) of the judges could contribute to the recommended performance standard on the Core Battery.

When the tests were considered as modules, the percent of items judged appropriate ranged from 92.3 for Form A of Professional Knowledge to 98.3 for Form B of Communication Skills. It is important to consider the item appropriateness of the modules because Core Battery scores are reported for the three modules rather than for their component tests.

TABLE 9

NUMBER AND PERCENT OF ITEMS RATED APPROPRIATE FOR EACH TEST FORM AND MODULE AND AVERAGE PERCENT APPROPRIATE FOR TEST FORMS AND MODULES

	FORM A		FORM B		AVERAGE[a]
	Number of Items	Percent of Items	Number of Items	Percent of Items	Percent of Items
Professional Knowledge Test/Module	96	92.3	96	94.1	93.2
Mathematics Test	24	96.0	25	100.0	98.0
Science Test	29	96.7	29	96.7	96.7
Social Studies Test	30	100.0	30	100.0	100.0
Literature/Fine Arts Test	31	88.6	27	79.4	84.0
General Knowledge Module	114	95.0	111	93.3	94.1
Reading Test	30	100.0	30	100.0	100.0
Listening Test	39	97.5	39	97.5	97.5
Writing Test (Objective)[b]	42	93.3	44	97.8	95.6
Communication Skills Module	111	96.5	113	98.3	97.4

[a] Sum of the percentages for Form A and Form B divided by 2.
[b] Excludes essay subtest.

Comprehensiveness of Tests

Judges were asked to identify any major topics that were included in the teacher education sequences of their institutions but did not appear in the content description of each test (see Appendix III-C). This was done in order to determine the extent to which the tests omitted important content. Forty-three of the 94 Content Review panelists completed a Test Content Omission Form. The results are summarized in Table 10.

Of the 43 panelists returning the form, 30 cited specific topics that were included in their institutions' teacher education programs but omitted in the Core Battery test content descriptions. The remaining 13 panelists who returned forms either indicated that no topics had been omitted or made general comments that were not content specific about the test or the teacher education curriculum. The number of omitted topics cited ranged from one (Reading Test) to 11 (Writing Test). Only seven topics were cited by two or more judges. Two of these topics were cited for the Listening Test: analysis and synthesis of oral communication, and stimulus-response questions for oral communication. Two judges noted that the Literature/Fine Arts Test did not include items on the recognition of names and works of famous artists. In the Mathematics Test, three panelists noted the omission of probability, and two cited statistics and algebraic problems as omitted topics. Finally, in the Writing Test (Objective), the topic of spelling of frequently confused words was reported to be omitted by two judges.

TABLE 10

TOPICS IN THE CURRICULUM IDENTIFIED BY JUDGES AS NOT INCLUDED IN THE TEST CONTENT DESCRIPTION OF THE NTE CORE BATTERY

am Section	Number of Judges Returning Form	Topic[a]	Frequency of Mention
ofessional Knowledge	6(3)[b]	Developmental processes	1
		Behavior modification	1
		Multicultural characteristics of contemporary school settings	1
		Pure psychology (as opposed to applied)	1
		Specific educational theorists & theories	1
thematics	8(1)[b]	Probability	3
		Statistics	2
		Algebraic problems	2
		Trigonometry	1
		Set theory	1
		Consumer oriented problems	1
		Informal geometry	1
		Properties of real numbers	1
ience	2	Motion	1
		Work	1
		Simple machines	1
		Physical energy	1
		Laboratory techniques	1
cial Studies	4(2)[b]	Contemporary American government	1
		Political issues and topics	1
		Constitutional and legal topics	1
		The Middle Ages	1
		European culture	1
		Early civilizations	1
		Evolutionary forces	1
iterature/Fine Arts	5(2)[b]	Recognition of names and works of famous people	2
		Growth and development in the arts	1
		Questions tied to content areas	1
eading	3(2)[b]	Comprehension of sequencing of events	1
istening	7(1)[b]	Analysis and synthesis of oral communication	3
		Stimulus-response questions for communication	2
		Audience analysis	1
		Rhetorical situations	1
		Modes of persuasion	1
		Psychology of listening vs. hearing	1
		Message strategies	1
		Logic of argument	1
		Minority cultures	1
		Following directions	1
		Evaluation of the message	1
riting	8(2)[b]	Spelling of frequently confused works (i.e., to-too, threw-through)	2
		Standard vs. non-standard English verb forms	1
		Transformational grammar	1
		Linguistic principles	1
		Rhetorical principles	1
		Comma splices	1
		Fused sentences	1
		Principal parts of verbs	1
		Library techniques	1
		Research papers	1
		Composition	1

---

er to Appendix III-C, Test Content Description, for a listing of topics included in the Core Battery Tests.

ber of judges within the total shown who returned form but indicated that no topics were omitted and/or wrote eral comments about the test.

Relative Emphasis of Tests and Curricula

The judges were asked to evaluate the match between the emphasis given content topics in the Core Battery tests and in Louisiana teacher education programs. Panelists were given a Test Content Description (Appendix III-C) showing the percentage of the test that was devoted to each content topic. They were then asked to judge whether that topic received the same, more, or less emphasis in Louisiana teacher education curricula than in the test. Panelists were told to disregard differences of 5 percent or less. This restriction was imposed to encourage judges to focus on only those meaningful differences that would indicate a true divergence between the tests and the curricula. Table 11 presents the frequencies of the three relative emphasis ratings for each test.

The column titled "Proportion of Same" in Table 11 shows the differences in emphasis for each test. The value given each test in this column was computed by dividing the number of "same" responses given by the panelists by the total number of responses ("same" plus "more" plus "less"). For example, the "Proportion of Same" value for the Professional Knowledge Test was:

$$\frac{\text{58 (number rating emphasis as "same")}}{\text{102 (number rating "same," "more," "less")}} = .57$$

When this proportion was .50 or greater it meant that, for the test as a whole, half or more of the panelists judged that the emphasis given different topics in teacher education programs in Louisiana was the same as the emphasis that topic received in the test.

TABLE 11

FREQUENCIES OF JUDGMENTS ABOUT RELATIVE EMPHASIS IN TESTS AND CURRICULUM

	Topic[a]	Percentage of Test	Number Rating Emphasis Given Topic in Curriculum:[b] Less	Same	More	Proportion of Same[c]	Degree of Difference[d]
ofessional Knowledge est/Module	1	24	2	11	4		
	2	25	3	10	4		
	3	17	5	10	2		
	4	9	7	7	3		
	5	11	6	10	1		
	6	14	6	10	1		
	Total	100	29	58	15	.57	-.32
Mathematics Test	1	20	2	6	3		
	2	16	2	7	2		
	3	24	1	7	3		
	4	16	2	9	0		
	5	12	4	7	0		
	6	12	4	7	0		
	Total	100	15	43	8	.65	-.30
Science Test	1	11	3	6	1		
	2	11	1	6	3		
	3	11	2	8	0		
	4	11	2	6	2		
	5	11	1	7	2		
	6	11	0	8	2		
	7	11	2	5	3		
	8	11	3	7	0		
	9	11	5	4	1		
	Total	99	19	57	14	.63	-.15
Social Studies Test	1	25	4	5	2		
	2	25	5	5	1		
	3	25	1	5	5		
	4	25	4	6	1		
	Total	100	14	21	9	.48	-.22
Literature/Fine Arts Test	1	29	1	5	4		
	2	54	5	4	1		
	3	17	3	4	3		
	Total	100	9	13	8	.43	-.06
neral Knowledge Module						.58	-.19
Reading Test	1	50	2	8	2		
	2	35	4	7	1		
	3	15	6	5	1		
	Total	100	12	20	4	.56	-.50
Listening Test	1	37	4	5	2		
	2	30	4	5	2		
	3	18	3	3	5		
	4	15	5	3	3		
	Total	100	16	16	12	.36	-.14
Writing Test (Objective)	1	55	4	6	1		
	2	45	4	5	2		
	Total	100	8	11	3	.50	-.45
mmunication Skill Module						.46	-.31

Refer to Appendix III-C, Test Content Description, for a description of topics included in the Core Battery Test.

Less: less emphasis in curriculum than test; Same: same emphasis in curriculum and test; More: more emphasis in curriculum than test.

Proportion of Same is total number of "Same" ratings divided by total number of all ratings.

Degree of Difference is algebraic sum of "Less" ratings (negative value) and "More" ratings (positive value) divided by total of "Less" and "More" ratings.

The "Degree of Difference" column provides an indication of the average or balance of the "more" or "less" ratings. The values in this column were computed by assigning algebraic signs to these responses (negative for "less," positive for "more") and then calculating the algebraic sum of these responses for each test. This sum was then divided by the total number of "more" and "less" responses for the test. For example, the "Degree of Difference" value for the Listening Test was:

Number responding "less":	16	Algebraic assignment:	-16
Number responding "more":	12	Algebraic assignment:	+12
Sum	28	Algebraic sum:	-4

$$\text{Degree of Difference: } \frac{-16+12}{16+12} = \frac{-4}{28} = -.14$$

A negative "Degree of Difference" meant that the judges were more likely to say that a topic was overrepresented in its emphasis on a Core Battery test compared to Louisiana teacher education programs. In other words, the topic received less emphasis in the curricula than in the test. A positive value indicated that, on the average, judges felt a topic was given less emphasis on the test than in teacher education curricula with which they were familiar.

As Table 11 shows, the "Proportion of Same" values ranged from .36 on the Listening Test to .65 on the Mathematics Test. For three tests the value was less than .50. That is, the number of panelists responding "more" or "less" outnumbered those responding "same." These tests were Literature/Fine Arts, Social Studies, and Listening. Additionally, the number of "less" responses exceeded the number of "more" responses for every test, as the negative "Degree of Difference" values indicate.

This meant that in cases in which panelists judged that the emphasis given topics in the Core Battery differed by more than 5 percent from the emphasis given those topics in teacher education curricula, the balance of these judgments was that the topics were overrepresented on the Core Battery tests. The values ranged from -.50 for Reading to -.06 for Literature/Fine Arts.

When the test ratings were averaged to give ratings for the modules they constituted, the "Proportion of Same" was greater than .50 for Professional Knowledge and General Knowledge. The value was .46 for Communication Skills. The "Degree of Difference" ranged from -.19 for General Knowledge to -.32 for Professional Knowledge.

Overall Similarity Between Core Battery Tests and Curricula

Panelists completed the Test Content Summary Form (Appendix III-D) to evaluate the overall similarity between the Core Battery tests and the teacher education sequences of their institutions. Each judge chose one of the following options to represent the degree to which the test he or she had reviewed paralleled the teacher education curriculum:

A. The test content topics parallel the teacher education sequence at our institution very closely.

B. There are some differences between the test content topics and the teacher education sequence at our institution, but these differences do not appear to be appreciable.

C. There appear to be some appreciable differences between the test content topics and the teacher education sequence at our institution.

D. There is little similarity between the test content topics and the teacher education sequence at our institution.

NR. No response was provided.

The response frequencies are shown in Table 12. The final column in the table, "Proportion Similar," was also used in the 1978 Validation Study to reflect the proportion of judges who indicated that the test they reviewed closely paralleled, or did not differ appreciably from, the curriculum. This number was computed by dividing the sum of the A and B responses for each test by the total number of responses (A plus B plus C plus D). The NR response was not included since it did not constitute a rating.

The results show that the values of "Proportion Similar" ranged from .40 in the Listening Test to .92 in the Reading Test. That is, 40 percent of the Listening Test panelists felt that there were no, or no appreciable, differences between the content of that test and the teacher education sequence; 92 percent of those evaluating the Reading Test made the same judgment. For the three modules the "Proportion Similar" ratings were .88 for Professional Knowledge, .80 for General Knowledge and .67 for Communication Skills. When all of the Core Battery tests were combined, 68 (76%) of the 89 judges who provided ratings felt that the content of the Core Battery tests closely paralleled or did not differ appreciably from that of the State's teacher education programs.

INTERPRETING AND EVALUATING CONTENT REVIEW RESULTS

The panelists provided four types of data about the content validity of the Core Battery tests. The first was concerned with the appropriateness of individual items within the tests. The other three types of data addressed the overall content validity of each test: the omis-

TABLE 12

FREQUENCIES OF JUDGMENTS ABOUT SIMILARITY BETWEEN THE CORE BATTERY TEST AND THE TEACHER EDUCATION SEQUENCE

	(A) Very Close	(B) No Appreciable Differences	(C) Some Differences	(D) Little Similarity	(NR) No Response	TOTAL	Proportion Similar[a]
Professional Knowledge Test/Module	3	11	2	0	1	17	.88
Mathematics Test	1	7	2	0	1	11	.80
Science Test	1	8	1	0	0	10	.90
Social Studies Test	0	9	2	0	0	11	.82
Literature/Fine Arts Test	0	6	3	0	1	10	.67
General Knowledge Module	2	30	8	0	2	42	.80
Reading Test	6	5	1	0	0	12	.92
Listening Test	1	3	5	1	1	11	.40
Writing Test (Objective)	0	7	4	0	1	12	.64
Communication Skills Module	7	15	10	1	2	35	.67
Total All Tests	12	56	20	1	5	94	.76

[a] Proportion Similar is sum of A + B divided by sum of A + B + C + D.

sion of important content topics, the match between the emphasis given content topics in the tests and in teacher education curricula, and the overall similarity between the content of the tests and teacher education programs. The 1978 validation study conducted by Educational Testing Service used each type of data as a criterion in evaluating the content validity of the tests reviewed at that time. A similar process was applied to the tests in the Core Battery for the current study.

Criteria for Special Review

A range of special review values was determined for each of the four criteria. Any test with observed values falling within the range of a special review value on any of the indicators of content validity was then reviewed in detail. Table 13 identifies the criteria used in judging whether a test qualified for special review. A description of the review criteria and the basis for the values assigned to each one are given below.

- Content appropriateness of items. An item was rated as content appropriate if 51 percent or more of the panelists reviewing the item judged that 90 percent or more of the students graduating from Louisiana teacher education programs would have had the opportunity to learn the knowledge required to answer the item correctly. Theoretically, the percentage of content appropriate items could range from 0 percent (none appropriate) to 100 percent (all appropriate). Following the procedures of the 1978 study, a test qualified for special review if less than 90 percent of the items in it were judged to be content appropriate.
- Comprehensiveness of test. The greater the number of topics cited by panelists as omitted from a test, the less comprehensive was the test. This number could have ranged from 0 (no topics omitted) to the number of content topics in a teacher education sequence (if the test included none of these). Following the procedures of the 1978 validation

TABLE 13

CRITERIA FOR PERFORMING SPECIAL REVIEW OF CORRESPONDENCE BETWEEN TEST CONTENT AND PROGRAM CONTENT

Aspect of Correspondence	Measure of Correspondence	Theoretical Range	Observed Range	Special Review Values
Content Appropriateness of Items	Percentage of questions classified as content appropriate by total total panel (average for two forms)	0% to 100%	84% to 100%	less than 90%
Comprehensiveness of Tests	Number of omitted content topics cited by two or more panelists	0 to total topics in curriculum	0 to 3	1 or more topics
Relative Emphasis of Tests and Curricula	a) Proportion of judgments rating emphasis as same	0 to 1.00	.36 to .65	less than .50 and
	b) Degree of Difference in "not same" judgments (- = curriculum emphasis "less", + = curriculum emphasis "more")	-1.00 to +1.00	-.06 to -.50	-.28 to -.50 (midpoint or below observed range)
Overall Similarity Between Tests and Curricula	Proportion of judges rating similarity as "closely parallels" or "does not differ appreciably"	0 to 1.00	.40 to .92	less than .50

study, the measure chosen for this criterion was citation by two or more judges. Citation by only a single judge was not considered representative of the total group evaluating the test, and requiring more than two citations could have excluded important omitted topics. A test qualified for special review if one or more topics were noted by two or more panelists as omitted.

- Relative emphasis. The first special review measure used here was the "Proportion of Same" judgments (that the emphasis given content topics within a test was the same as the emphasis given those topics within teacher education curricula). This value could range from 0 (emphasis differed for all topics) to 1.00 (emphasis the same for all topics). It should be noted that judges gave a rating of "same" only when they felt that the content emphases of the test and the curriculum were within 5 percent of one another. Hence, this is a conservative measure of similarity. There was no standard value for what constituted acceptable "Proportion of Same" ratings; .50 was chosen as a plausible value because below this level a test became more dissimilar than similar in content emphasis to teacher education curricula.

  A second measure was also used in identifying tests that were candidates for special review because of differences in relative emphasis. The "Degree of Difference" value indicated whether panelists felt that a test differed in emphasis from the curriculum by providing less emphasis (positive value) than the curriculum or more emphasis (negative value). This value could range from -1.00 (all non-same topics overrepresented on test) to +1.00 (all topics underrepresented on test). A test could qualify for special review if its "Degree of Difference" value was at or above the midpoint of the observed values for all tests. Since the observed values ranged from -.06 to -.50, the special review value was set at -.28. This figure was computed by adding -.06 and -.50 and dividing the sum by 2. Tests with a value of -.28 to -.50 were potentially eligible for special review. This process identified tests with a greater than average overemphasis of content topic. Using two factors together--the "Proportion of Same" judgments and the size of the "Degree of Difference"--provided joint estimates of relative emphasis. Therefore, tests were considered eligible for special review only if they fell within the specified value ranges for both criteria.

- Overall similarity. This measure was the proportion of judges who felt that a test closely paralleled or did not differ significantly from the curriculum. It could range from 0 (all judges felt that the test differed appreciably from, or bore little similarity to, the curriculum) to 1.00 (all judged that the test closely paralleled, or did not

> differ appreciably from, the curriculum). Tests were selected for special review if this value was less than .50; that is, if fewer than half of the panelists judged that the overall similarity between the test and the curricula was high. The same criterion value had been employed in the 1978 study.

Special Review Criteria Results

The observed value for each of these criteria was compared with the special review value for the eight tests comprising the Core Battery. Table 14 presents the results. Four tests (Professional Knowledge, Science, Social Studies, and Reading) had values within the accepted ranges on all criteria. None of these tests warranted special review.

The Literature/Fine Arts Test had an item appropriateness value of 84.0 percent. It was the only test requiring further review for item appropriateness. Although the number of items was not large enough to support statistical testing of distribution, five of the 10 items judged inappropriate concerned a single topic. This was the topic of "relating works of literature and fine arts to one another and their social-historical context." It constituted 17 percent of the test.

Mathematics, Literature/Fine Arts, Listening, and the objective subtest of the Writing Test qualified for special review on the basis of comprehensiveness, since each test omitted one or more topics cited by two or more judges. None of the tests fell within the range of review values for both factors in the relative emphasis criterion. The Listening Test was the only one of the eight tests to qualify for special review on the criterion of overall similarity.

TABLE 14

SPECIAL REVIEW CRITERIA DATA FOR CORE BATTERY TESTS

	Appropriateness (< 90%)	Comprehensiveness (≥ 1 topic omitted)	Relative Emphasis (< .50)	(-.28 to -.50)	Overall Similarity (< .50)
Professional Knowledge	92.9	0	.57	-.32*	.88
Mathematics	98.0	3*	.65	-.30*	.80
Science	96.7	0	.63	-.15	.90
Social Studies	100.0	0	.48*	-.22	.82
Literature/Fine Arts	82.9*	1*	.43*	-.06	.67
Reading	100.0	0	.56	-.50*	-.92
Listening	97.5	2*	.36*	-.14	.40*
Writing (Objective)	95.6	1*	.50	-.45*	.64

* Test fell within range for special review on this criterion

Application of Special Review Criteria Results

The impact of the review criteria lay in each test's contribution to the content validity of the Core Battery module of which it was a component. This was because scores are reported for entire modules rather than individual tests or the Core Battery as a whole. Therefore, in evaluating the impact of the special review criteria, it was necessary to assign a numerical weight to each test that reflected the number of criteria on which the test qualified for special review, and the test's contribution to the total module. This was a two-step process.

The first step paralleled the procedure used in the 1978 validation study. Each test was assigned a value based upon the number of special review criteria for which it qualified. These values and their interpretations were:

0: No criteria for special review, test very closely related to curricula.
1: One criterion for special review, test closely related to curricula.
2: Two criteria for special review, test reasonably related to curricula.
3: Three criteria for special review, test probably related to curricula.
4: Four criteria for special review, test not related to curricula.

The second step was to multiply the value given each test by the number of items in that test, and to sum the resulting figures for all tests within a module. The results are shown in the "Items x Value" column of Table 15. This sum was then divided by the total number of items in a module to produce the Module Value weights shown in the Table. In this way the content validity results for each test contributed as much in judging the content validity of the module as performance on the test would contribute to the module score.

TABLE 15

CONCLUSIONS ABOUT CORRESPONDENCE BETWEEN CORE BATTERY MODULES AND PROGRAM CONTENT

	Mean Number of Items (Form A and Form B)	Special Review Value	Items x Value	Module Value[a]	Interpretation
Professional Knowledge	103	0	0	0.000	Very Closely Related
General Knowledge					
Mathematics	25	1	25		
Science	30	0	0		
Social Studies	30	0	0		
Literature/Fine Arts	34.5	2	69		
TOTAL	119.5		94	0.787	Closely Related
Communication Skills					
Reading	30	0	0		
Listening	40	2	80		
Writing (Objective)	45	1	45		
TOTAL	115		125	1.087	Closely Related

---

[a] Module Value equals Items x Value divided by Number of Items.

Both the Listening and Literature/Fine Arts Tests had a weight of 2 (reasonably related). For Mathematics and Writing (objective subtest) the weight was 1 (closely related). The other four tests all had weights of 0, or very closely related.

When these weights were adjusted to account for the contribution of each test to the total score of its respective module, the modules demonstrated acceptable levels of content validity. Professional Knowledge was the only test within its module, and it produced a module weight of 0.0 or "very closely related" to the teacher education sequence in Louisiana. The other two Core Battery modules, General Knowledge and Communication Skills, had weights of about 0.8 and 1.1, respectively. This was close to the value of 1.0 that was interpreted as "closely related" to the State's teacher education curricula.

These interpretations are not meant to discount content differences between Louisiana's teacher preparation programs and the individual Core Battery tests. They simply provide another way of evaluating the content validity of the tests and, by extension, the total modules. It should be stressed that the methodology and values used in the special review criteria are not absolutes, but were adapted from the precedent of earlier validation studies. The special review information, therefore, is presented to the Blue Ribbon Score Committee and to the Superintendent of Education for their use in making informed judgments about the content validity of the Core Battery.

CONSISTENCY OF RESULTS

The 1978 validation study reported the consistency of ratings across half-panels for the content appropriateness of items and the relative emphasis of tests and curricula. After careful deliberation, the investigators decided to follow the same procedure in the current study and to provide consistency information for content appropriateness and relative emphasis. These constituted two major rating tasks for the Content Review Panels, and these tasks were considered sufficient for half-panel contrasts.

Consistency of Results for Content Appropriateness of Items

The consistency of judgments about the content appropriateness of the items was examined from two points of view. Table 9 reports the extent to which the percentages of content appropriate questions were similar for the two forms of each test. The extent to which the half-panels agreed on the item appropriateness of test forms is shown on Table 16. As in other tables, the half-panels are identified for convenience as Panel 1 and Panel 2.

Table 16 shows the percentage of items judged content appropriate by half-panel and test form, as well as the average of Forms A and B. The half-panel average ratings differed by 10 or more percentage points for only two tests. These were Professional Knowledge, in which the average ratings by the half-panels differed by 14.6 percentage points, and Listening, in which the difference between half-panels was 10.0 points. For two tests, Mathematics and Reading, the average percentages

TABLE 16

NUMBER AND PERCENT OF QUESTIONS JUDGED CONTENT APPROPRIATE, BY TEST FORM AND HALF-PANEL

	Panel 1					Panel 2				
	Form A		Form B			Form A		Form B		
	Number Of Items	Percent Of Items	Number Of Items	Percent Of Items	Average Percent Of Items	Number Of Items	Percent Of Items	Number Of Items	Percent Of Items	Average Percent Of Items
Professional Knowledge	98	94.2	97	95.1	94.7	82	78.8	83	81.4	80.1
Mathematics	23	92.0	25	100.0	96.0	24	96.0	23	92.0	94.0
Science	30	100.0	30	100.0	100.0	28	93.3	28	93.3	93.3
Social Studies	30	100.0	30	100.0	100.0	26	86.7	29	96.7	91.7
Literature/Fine Arts	29	82.9	27	79.4	81.1	28	80.0	24	70.6	75.3
Reading	30	100.0	30	100.0	100.0	30	100.0	28	93.3	96.7
Listening	33	82.5	37	92.5	87.5	39	97.5	39	97.5	97.5
Writing (Objective)	41	91.1	42	93.3	92.2	38	84.4	40	88.9	86.7

across test forms differed by four or fewer points. These results are similar to those reported in the 1978 validation study.

Table 17 provides information about the extent to which the half-panels agreed on the classification of individual items. The columns in the Table report the percentage of items that both half-panels felt were appropriate, that the half-panels disagreed upon, and that both half-panels felt were inappropriate. Finally, the columns identified as "Percentage of Agreement" contain the percentage of items that both half-panels agreed upon as either appropriate or inappropriate.

The percentages of agreement ranged from 74.3 (Literature/Fine Arts) to 100.0 (Reading) on Test Form A. The range was smaller for Form B: from 82.2 (Writing) to 96.7 (Social Studies). Of the 16 percentages of agreement calculated, 14 were greater than 84 percent. By contrast, only 29 of the 46 percentages of agreement calculated for the Common and Area Examinations in the 1978 validation study exceeded 84 percent.

An average probability of agreement was also calculated. This probability was estimated as the average of two conditional probabilities that are given by: a) the percentage of items judged content appropriate by both Panels 1 and 2 divided by the percentage of items judged appropriate by Panel 2, and b) the percentage of items judged appropriate by Panels 1 and 2, divided by the percentage of items judged appropriate by Panel 1. This statistic is a measure of the probability that both of the half-panels would judge an item appropriate. The example below shows how the average probabilities of agreement were calculated for the Listening Test, averaged across test forms.

TABLE 17

AGREEMENT BETWEEN HALF-PANELS ABOUT
CONTENT APPROPRIATENESS OF ITEMS, BY TEST FORM

	Form A				Form B			
	Both Panels Appropriate	Panels Disagree	Both Panels Inappropriate	Percentage of Agreement	Both Panels Appropriate	Panels Disagree	Both Panels Inappropriate	Percentage of Agreement
Professional Knowledge	78.8	15.4	5.8	84.6	81.4	13.7	4.9	86.3
Mathematics	88.0	12.0	0.0	88.0	92.0	8.0	0.0	92.0
Science	93.3	6.7	0.0	93.3	93.3	6.7	0.0	93.3
Social Studies	86.7	13.3	0.0	86.7	96.7	3.3	0.0	96.7
Literature/Fine Arts	68.6	25.7	5.7	74.3	67.6	14.7	17.6	85.2
Reading	100.0	0.0	0.0	100.0	93.3	6.7	0.0	93.3
Listening	82.5	15.0	2.5	85.0	90.0	10.0	0.0	90.0
Writing (Objective)	80.0	15.6	4.4	84.4	84.4	17.8	0.0	82.2

Percent of Items Judged by:
Panel 1:

Panel 2:	Appropriate	Inappropriate	Total
Appropriate	86.25	11.25	97.50
Inappropriate	1.25	1.25	2.50
Total	87.50	12.50	100.00

From this table, the conditional probability of agreement for items judged content appropriate was computed as:

$$\frac{\frac{.8625}{.9750} + \frac{.8625}{.8750}}{2} = .936$$

In a similar fashion, half-panel consistency in agreement about items judged inappropriate was viewed as the probability that one of the panels would judge an item inappropriate, if the other panel had judged it so. For the Listening Test, this probability was:

$$\frac{\frac{.0125}{.0250} + \frac{.0125}{.1250}}{2} = .300$$

The conditional probabilities of agreement for all of the Core Battery Tests are presented in Table 18. The values for the items classified content appropriate were consistently high, ranging from .87 to .98. In other words, if the judges on one half-panel determined that an item was appropriate, the judges on the other half-panel generally agreed.

On the other hand, there was less agreement about items that were judged inappropriate. The values ranged from .00 for three tests (Mathematics, Science, and Social Studies) to .63 (Professional Knowledge). Only three of the eight tests had an average probability of agreement on

TABLE 18

AVERAGE PROBABILITY OF AGREEMENT ON ITEM APPROPRIATENESS ACROSS HALF-PANELS[a]

	Appropriate	Inappropriate
Professional Knowledge	.92	.63
Mathematics	.95	.00
Science	.97	.00
Social Studies	.96	.00
Literature/Fine Arts	.87	.55
Reading	.98	.50
Listening	.94	.30
Writing (Objective)	.91	.23

[a] Note that the data used to compute these figures are based on the average of content appropriate and inappropriate percentage responses for Forms A and B. Refer to text for an example of the calculations.

inappropriateness that was .50 or greater. These low values point to one of the difficulties that arise in constructing a test to measure knowledge of the content among different teacher preparatory programs. Most items will match the content of most curricula; others will not; but not all test items will match all the topics of all of the curricula.

It is interesting to note that the tests on which there was .00 agreement with regard to content inappropriate items (Mathematics, Science, and Social Studies) had very high probabilities of agreement on appropriate ratings (.95, .97, and .96 respectively). The panels for these tests were composed of members from the most diverse subject areas, who apparently agreed on appropriate items, but disagreed greatly on inappropriate items. The disagreement on inappropriate items very likely reflects the diversity in curricula among the panelists' universities and departments.

High average probability of agreement for inappropriate items occurred for the Professional Knowledge, Literature/Fine Arts, and Reading Tests (.63, .55, and .50 respectively). This suggests that there was closer agreement about the content of these fields across the universities and the departments represented by the half-panel members for these tests than among other panelists.

Consistency of Results for Relative Emphasis

As reported earlier in Table 10, the analysis of the relative emphasis of tests and curricula focused on two values: (a) the "Proportion of Same," which indicated for each test as a whole the percentage of panelists who felt the topics received the same emphasis in the teacher education curriculum as in the test; and (b) the "Degree of Difference," which indicated the average balance of the "more" or "less" ratings. The half-panel analyses shown on Table 19 compare the half-panel ratings on these same two criteria.

The "Proportion of Same" ratings show a high agreement between the half-panels (differences in proportion of 7 percent or less) for four tests: Mathematics, Science, Social Studies, and Listening. The half-panels for these examinations agreed on the proportion of topics that received the same emphasis on the test as in their teacher education curricula. There was somewhat less agreement between the half-panels on the "Proportion of Same" ratings for the other four examinations, ranging from a difference of 11 percent for Reading to a difference of 28 percent for Professional Knowledge.

The results for "Degree of Difference" indicate that the half-panels agreed in directionality for five of the eight tests. These were the Reading, Writing, Literature/Fine Arts, Mathematics, and Professional Knowledge Tests, in which all half-panel ratings were negative except for a .00 rating produced from one of the half-panels in Literature/Fine Arts. Both half-panels agreed, on the average, that differences in emphasis reflected an overrepresentation of content topics on these tests.

There were disagreements in "Degree of Difference" between the half-panels on the Listening, Science, and Social Studies Tests. It is interesting that these all had high half-panel agreement on the "Propor-

TABLE 19

HALF-PANEL RATINGS FOR RELATIVE EMPHASIS OF TESTS AND CURRICULA[a]

	Panel 1		Panel 2	
	Proportion of Same	Degree of Difference	Proportion of Same	Degree of Difference
Professional Knowledge	.70	-.13	.42	-.43
Mathematics	.64	-.38	.67	-.20
Science	.64	.13	.62	-.41
Social Studies	.50	.40	.46	-.69
Literature/Fine Arts	.53	-.14	.33	.00
Reading	.61	-.71	.50	-.33
Listening*	.33	-.38	.40	.17
Writing (Objective)	.40	-.33	.58	-.60

[a] These data were generated in the Fall validation procedure by persons who reviewed only Form A of the Listening Test. The half-panels for this analysis were those constructed for Form A of the test. See Chapter II for an explanation of assignment to half-panels for the Listening Test.

tion of Same" ratings. It appears that the half-panels for these three tests agreed highly on the proportion of topics that received the same emphasis in their college curricula, but were divided in their opinions of whether those topics that were disproportionately represented received "more" or "less" emphasis in their curricula.

On the other hand, half-panels with relatively higher disagreement on the "Proportion of Same" ratings (Reading, Writing, Literature/Fine Arts, and Professional Knowledge Tests) agreed on the directionality of difference for those topics that were disproportionately represented. It appears that high agreement between half-panels occurred on either "Proportion of Same" ratings or on "Degree of Difference" ratings, but not on both. The half-panels for the Mathematics Test were the only ones to agree highly on both "Proportion of Same" and "Degree of Difference."

SUMMARY

The results presented in this chapter suggest that the Content Review Panels generally evaluated the Core Battery tests as appropriate measures of knowledge for teacher education graduates in Louisiana. For all tests but Literature/Fine Arts the panelists rated 90 percent or more of the items as appropriate. _It should be remembered that items judged inappropriate by half or more of the panelists reviewing them will not be included in calculating the Knowledge Estimation Panel's recommended performance level for the Core Battery_. In four of the eight tests two or more judges cited a topic as omitted. However, in these tests the number of omitted topics was still small, ranging from one topic (Writing and Literature/Fine Arts) to three (Mathematics). The "Proportion of Same" ratings show that for the majority of the tests (five of eight) the panelists also judged that the emphasis placed on content topics in their teacher education curricula was acceptably close to the emphasis given these topics in the Core Battery.

The numerical weighting used to convert the panelists' judgments on a variety of content validity measures to a single rating of each module's overall appropriateness (see Table 15) demonstrates the acceptability of the Core Battery. The Professional Knowledge Module was rated as "Very Closely Related" to the Louisiana teacher education curricula. Both the General Knowledge and Communication Skills Modules were evaluated as "Closely Related."

CHAPTER IV

KNOWLEDGE ESTIMATION RESULTS

INTRODUCTION

The purpose of this chapter is to present the analyses of the judgements of the Knowledge Estimation Panel that yielded information for determining the performance standard on the NTE Core Battery. The chapter is divided into four major sections. The first section contains a discussion of standard setting and standard setting methods. The second section contains an overview of the knowledge estimation process used in this study. Included are an explanation of the purpose of the Knowledge Estimation Panels, a review of the training provided to judges, a description of the tasks performed, and an explanation of the relationships among the various standards that resulted from the Knowledge Estimation ratings. The third section of the chapter contains the results of the panel ratings of the test items. Half-panel results for the Knowledge Estimation Panels are reported in the final section of the chapter.

INTRODUCTION TO STANDARD SETTING

The use of tests to classify examinees into categories or states based on test performance has increased in the last 15 years. Tests are being used to make decisions ranging from the formation of classroom groups to the certification and licensure of individuals desiring entrance into various professions. The use of the NTE for certification is an example of the latter use of tests. One feature that must be present when tests are used to make classification decisions about examinees is a score that represents the decision point (the point that is used to separate acceptable from unacceptable candidates). The issues of how to establish qualifying scores and the validity of the resulting classifications have been the subject of much debate in the psychometric community.

Numerous methods have been proposed for setting standards. Descriptions of many of the methods have been compiled by several researchers (e.g. Hambleton & Eignor, 1979; Jaeger, 1976; Popham, 1978; Shepard, 1976). One common feature of all standard setting methods is that they employ human judgment. All standard setting methods, even the ones that use test data must, at some point, involve subjective judgments about the final placement of the standard. This is not, however, a necessarily undesirable condition. Many useful and accepted standards in areas other than measurement are subjective. Speed limits, acceptable discharge limits of chemicals from industrial plants, and numerous other standards that affect daily life are the result of subjective judgment. Since totally objective criteria to determine standards are generally unavailable, it is important to ensure that the individuals whose judgments are solicited are qualified to make the judgments required of them. Such is the case with the determination of qualifying scores on the NTE Core Battery. The test items were reviewed by professors who were nominated by their respective deans as being qualified to perform the ratings. Likewise, the members of the Blue Ribbon Score Committee were selected for the contributions their respective backgrounds and experience could bring to the task.

Hambleton and Eignor (1980) have classified a number of standard setting methods into three categories: Judgmental, Empirical, and Combination. Empirical and Combination methods require actual test data and, in most cases, some external measure of the examinee's status to determine the placement of the standard. For example, the Contrasting Groups method (a Combination method) requires that judges (usually instructors) rate each examinee as above or below the minimally acceptable performance level. Test score distributions of the two groups can be compared. The point of intersection of the two distributions is a logical choice for a performance standard.

Such data were not available for the examinee population who would take the NTE Core Battery for teacher certification. Therefore, one of the Judgmental methods of setting standards was selected. These methods require judges to rate test items to establish a qualifying score. The three most commonly mentioned Judgmental methods are the Nedelsky (1954), Angoff (1971), and Ebel (1972) methods. Each method is briefly described below.

Nedelsky

The Nedelsky method requires judges to determine the attractiveness to the minimally knowledgeable examinee of each distractor of a multiple choice test item. Judges read the test item and the options presented with it. Each judge then eliminates the options he or she feels would be eliminated as clearly incorrect by the minimally knowledgeable examinee. It is assumed that the minimally knowledgeable student will eliminate options he or she feels are clearly wrong and then guess randomly among the remaining options.

Thus, the probability of the minimally knowledgeable person's correctly answering an item becomes the reciprocal of the number of options remaining after clearly incorrect ones have been eliminated. For example, for a five choice item, a judge may feel that the minimally knowledgeable student would eliminate two options as clearly incorrect. Since three options remain, the estimated probability of a correct answer is one third (1/3).

One feature of the Nedelsky method that can be problematic is that judges' ratings result in fixed and unequally spaced probabilities. If all incorrect options are eliminated by a judge, the probability of a correct response is 1.0. If two options remain, the probability is 0.50. Under no circumstances may the probability of a correct response be 0.75 or any other number between 0.50 and 1.00. This restriction in the method has been regarded as a drawback. However, the Nedelsky method does require judges to rate each distractor. Judges must consider the fine distinctions between distractors that may make them differentially attractive to examinees. This level of scrutiny is not found in other standard setting methods.

Angoff

The Angoff method also requires judges to consider test items from the frame of reference of the minimally knowledgeable examinee. In the Angoff method, however, the judge is required only to estimate the probability that the minimally knowledgeable student would answer the item correctly. In effect, this method requires judges to predict the item difficulty of each item for the minimally knowledgeable group of examinees. The sum of these probabilities equals the predicted test score for the minimally knowlegeable examinee.

The major criticism of the Angoff method is the judges' inability to predict accurately item scores for examinees (Livingston & Zieky, 1982). This is an unfamiliar task to most judges, and studies to date have not indicated that judges can do it accurately. While judges often work with test scores and can often predict the overall test performance of an examinee, it is much more difficult to predict performance on a specific item.

A modification of the Angoff method was used in the 1978 NTE validation study in Louisiana and in other studies. Judges were provided with a seven-point scale (2, 10, 25, 40, 60, 80, and 98) and were instructed to choose one of the points to represent the predicted probability of a correct answer. Also included in the scale was an "I Don't Know" option. This method still requires judges to predict item difficulties, but the scale limits the options available and restricts the predicting of item difficulty. Additionally, some researchers maintain that the presentation of a fixed scale may influence judges to make certain types of ratings. For example, on questionnaires there is often a tendency among respondents to choose options toward the middle of the scale. Similar results may occur with the modified Angoff.

Ebel

The Ebel (1972) method is a more complex method than the Nedelsky or Angoff methods. It involves rating each test item on two dimensions: relevance and difficulty. Ebel's original presentation of the method used four relevance categories (essential, important, acceptable, and questionable) and three difficulty levels (easy, moderate, and hard). However, the number of categories within a dimension is not prescribed. Once each item has been rated along each of the two dimensions, judges assign a percentage to each cell formed by a combination of a relevance and a difficulty category. The percentage is the judge's expectation of the performance of the minimally knowledgeable examinee on the items in that combination.

The Ebel method was selected for use in this study for several reasons. First, the Department of Education's researchers felt that limiting the design of the study ratings of item difficulty to three broad categories (easy, moderate, and hard) would yield more reliable and valid data than would the Angoff seven-point rating scale. Second, there were no restrictions on the assignment of difficulty predictions as in the Nedlsky and modified Angoff methods. Judges were free to choose any percentage from 0 to 100 for each category. Third, the task of estimating the proportion of correct scores for a group of similar items is not unlike estimating a total test percent score. Since this is a level of evaluation more familiar to judges, it was felt that they would provide more valid information with that task than with that of rating individual items. Finally, the addition of a relevance dimension to the difficulty dimension (the only dimension rated in the 1978 Educational Testing Service study) appeared to be a significant enhancement of the previous validation methodology. Both item difficulty and the relevance of items were included in the calculation of the standard.

Description of the Ebel Methodology Employed

A nine-cell Ebel method was used for the current NTE Core Battery study. Items were rated on the three difficulty dimensions described previously and on three relevance dimensions (essential, important, not very important). The resulting 3 x 3 matrix is displayed in Figure 3.

	Item Difficulty		
Item Relevance	Easy	Moderate	Hard
Essential			
Important			
Not Very Important			

FIGURE 3: EBEL DECISION MATRIX

Judges were presented with two test forms corresponding to the fall, 1982, and spring, 1983, editions of the Core Battery. Each item was shown with the correct answer indicated. The judges also received a rating form with two columns (see Appendix IV-A). For each item, a judge was instructed to circle an E, M, or H (easy, moderate, hard) in the difficulty column to reflect his or her difficulty rating, and an E, I, or N (essential, important, not very important) in the relevance column to represent the relevance rating. After the difficulty and relevance of all items were rated, a second form was completed by the judges. The second form was used to indicate the percent of items that the judge expected a minimally knowledgeable teacher candidate to answer correctly in each difficulty/relevance category (e.g., easy/essential, moderate/important, etc.). A rating of 0 percent would mean that a minimally knowledgeable examinee would answer none of the items in the category correctly, whereas a 100 percent rating would mean that all items in the category would be answered correctly.

Description of the Essay Methodology

While the Ebel method seemed well-suited to the rating of the objective items included in the Core Battery, it was clear that a different method would have to be employed for the essay subtest of the Writing Test. In this case, it did not seem feasible to ask judges to read the test item alone and to estimate a minimally knowledgeable essay response. It was also not possible for judges to determine an acceptable score without seeing scored essays. For that reason, it was determined that judges would require access to actual essays written by examinees taking the Core Battery in order to determine what constituted minimal writing performance (pilot test essays were not available).

Following the fall, 1982, administration and scoring of the Core Battery, a sample of 60 essays was drawn from the set of essays written by Louisiana examinees and scored by Educational Testing Service. Essays were scored by Educational Testing Service using a holistic scoring methodology.

Each essay received a score between 0 and 12, inclusive. Since the scoring methodology was not part of the validation effort, it is not discussed here. The sample was drawn to approximate the distribution of essay scores in the State. The writing panelists returned to Baton Rouge in February, 1983, to read the essays. Both Content Review and Knowledge Estimation panelists rated the essays, assigning each a rating of acceptable, unacceptable, or borderline. The average score assigned to those papers rated as borderline was determined for each judge. The mean of these averages was taken as the recommended standard on the essay subtest.

KNOWLEDGE ESTIMATION PROCESS

Purpose of Knowledge Estimation Panels

The NTE serves as a certification instrument in Louisiana. Individuals wishing to teach in Louisiana must, in addition to completing a teacher education program, take the NTE and reach or exceed established performance levels. For most prospective teachers, this has meant taking both the Common Examinations and a test specific to each candidate's major field (an Area Examination). A combined Area and Weighted Common Examinations Total score has been used to determine certification. With the replacement of the Common Examinations by the Core Battery, existing certification standards are no longer applicable. Instead, a combination of the Core Battery module scores and an Area score will be used in the certification decision. (See Chapter V for a discussion of the format of qualifying scores.)

Since the Core Battery scores are to be used in the certification decision, it is necessary to establish a performance standard. This standard will represent the score at which an individual is judged to have sufficient knowledge to begin a teaching career in Louisiana. The Knowledge Estimation panels were formed to provide information to assist in determining that standard.

Panelists were selected to evaluate items in one of eight subject areas of the tests (Professional Knowledge, Mathematics, Science, Social Studies, Literature/Fine Arts, Reading, Listening, and Writing). Each participating panelist reviewed test items from the content area to which he or she was nominated. A description of the selection of panelists can be found in Chapter II of this report. While Knowledge Estimation judges performed only one conceptual task, there were two major steps in that task.

Formulation of a Reference Group

Members of the Knowledge Estimation panels were required to develop a hypothetical reference group in order to make specific judgments about the knowledge possessed by members of that group.

First, the judges were asked to refer primarily to those graduates who were likely to become teachers. If judges were familiar with their students' career plans, they were instructed to refer only to students who were planning to pursue teaching careers. Panelists were directed to refer to graduates in all teacher education fields who intended to teach in the elementary or secondary grades.

Second, judges were asked to refer only to those graduates who were just minimally knowledgeable. "Minimally knowledgeable" was defined as having the minimal amount of academic knowledge to a) complete the college program required for certification in the State and b) teach effectively. This task was closely related to the practical experiences of the panel members and thus was likely to be performed with reasonable accuracy. An assessment of minimal knowledge was a task that every faculty member performed, at least in part, every time he or she wrote and graded examinations for students enrolled in college courses or evaluated a student teacher's performance. The dividing line between a minimally passing and a failing grade must be established by a faculty member in designing an examination to preclude setting the level of difficulty of the questions so high that it excludes the minimally qualified student from demonstrating the level of knowledge he or she has. That dividing line must be assessed again in grading examinations; because the delivery of a failing grade has such important consequences, most conscientious faculty members pay great attention to their conception of what the minimally qualified student should be able to do to achieve a passing grade on their examinations. The Knowledge Estimation Panel members were asked to draw upon this knowledge in setting the dividing line between low passing and failing work, and to apply their conception of the minimally knowledgeable student to test questions written by others.

Since the standard setting process was a difficult and unfamiliar one to most Knowledge Estimation panelists, a half-day training session preceded the rating task. The purpose of the training session was to familiarize judges with the tasks to be performed and to help each judge define for himself or herself the "minimally knowledgeable" teacher candidate. A description of that training is provided later in the chapter.

Formulation of the Estimates

In the development of the study design, it was anticipated that many faculty members would not have taught all of the courses in which the content of the test items would have been covered. The Social Studies Test, for example, contains content from six separate disciplines. Prior to panel assemblies, the judges were sent material that encouraged them to consult with local sources of information regarding the curriculum at the institution in which they taught, such as college catalogs, specialists in curriculum planning, or other available sources before attending the panel assemblies. They were also told that they might wish to talk with colleagues who had taught specific courses that they themselves had not been called upon to teach. An overview of the panelists' tasks is included in Appendix IV-B.

Judges were instructed to base their predictions of percentage scores in each difficulty/relevance cell on the probability of correct responses by the minimally knowledgeable teacher candidate. In the 1978 validation study of the NTE in Louisiana, judges had been asked to make estimates about knowledge rather than correct responses. This was done to free judges from the responsibility of attempting to determine the extent to which minimally knowledgeable students would guess the correct answer. Since Common Examination scores were corrected for guessing, this was an appropriate procedure. Scores from the Core Battery, however, are not corrected for guessing. Therefore, judges participating in the current study were directed to make estimates concerning the percentage of items the minimally knowledgeable teacher candidate would answer correctly.

The judges were asked to make estimates for all items except those for which their experience provided them with no basis for making a judgment. In these cases they were instructed not to respond to that item.

Knowledge Estimation Training Session

In order to familiarize judges with the rating task they were to perform, a training session was held the afternoon before the actual rating session was to occur. The training session was organized into three meetings.

First, all Knowledge Estimation Panel members convened in a single large-group session. General issues were addressed and an overview of the tasks to be performed was presented in this meeting.

Following the general session, Mixed Content (MC) mini-sessions were held. Each of the eight MC mini-sessions contained panelists from a variety of subject areas. The purpose of the MC sessions was to stimulate discussion of the concept of the "minimally knowledgeable teacher candidate." Each panelist received a copy of a handout describing the minimally knowledgeable teacher candidate (Appendix IV-C). The panelists then participated in a discussion of their interpretations of the description in order to clarify each judge's perception of the minimally knowledgeable teacher candidate. The discussions covered curriculum issues and broad skill areas in addition to the description of the minimally knowledgeable teacher candidate provided in the handout. Panelists were told that the purpose of the discussion was not to reach consensus on what constituted minimal knowledge, but rather to aid each panelist in clarifying and stabilizing his or her own conception of the minimally knowledgeable teacher candidate. In other words, each panelist made a set of independent judgments. Before doing so, however, a training forum was provided to develop a stable concept of the minimally knowledgeable teacher candidate in each judge's mind. The MC sessions lasted approximately one hour.

Following the MC mini-sessions, Same Content (SC) sessions were held. Participants in each SC session were nominated to serve in evaluating the same Core Battery Test. SC sessions began with a brief review of the MC sessions. Following the initial discussion, judges were provided with sample item sets, forms, and directions for completing the Knowledge Estimation rating task (Appendixes IV-D and IV-E contain sample forms and directions). Seven of the panels were provided with 10 sample items. The Professional Knowledge Panel received 20 sample items since their actual rating task was to be significantly longer than that of the others panels. Half of the items were rated, and these ratings were tabulated and discussed. Then the remaining half of the items were rated, tabulated, and discussed. Judges then predicted the proportion of items in each category that would be answered correctly by the minimally knowledgeable teacher candidate. Following the rating of the sample items, panelists entered their results into a knowledge estimation training worksheet (Appendix IV-F) to demonstrate how the ratings would later be used to determine qualifying scores. These results were also compiled and discussed.

Description of Ratings

Item ratings were conducted on the morning following the training session. Ratings were counterbalanced by both form and task. That is, half of the judges rated Form B first, and the others rated Form A first; half rated item difficulty first, and half rated item relevance first. Items that were common to both forms were deleted from the Form B packets. During data processing, ratings of common items (those that appeared on both Form A and

Form B) were included in the analysis of each test form to develop comparable ratings. A large general session was held for the rating task. Panelists evaluating the same test were encouraged not to sit with one another, and judges were instructed not to confer among themselves.

At the time of the ratings, only one form of the Listening Test was available in final form. While a typed version of the final form was available, it was felt that judges should hear the Listening Test rather than read it. For that reason, only the one complete form of the Listening Test was rated in the October, 1982, session. Panelists were invited to return to Baton Rouge in February, 1983, to rate the second Listening Test form. Seven judges returned for this rating session. At that time, they listened to the first form of the test and reviewed their ratings of that form to re-establish the framework from which they made their Form A ratings. Following the review, each judge rated Form B in the same manner in which he or she had rated the first form.

Since the essay validation methodology required a sample of examinees' essays, ratings of the essay subtest could not be conducted during the large rating session in October. Instead, Writing panelists reconvened in February, 1983, to rate the essays. A sample of 60 essays produced by Louisiana examinees was chosen to approximate the overall distribution of essays in the State. Panelists from both the Content Review and the Knowledge Estimation Writing Panels were invited to participate in the ratings. Only essays from the fall, 1982, administration of the Core Battery were available for rating. The time constraints for completion of the study did not allow for rating the spring, 1983, essay form.

Relationship of Panel Judgments to Module Qualifying Scores

As explained previously, eight panels of judges (one for each test in the Core Battery) were formed to provide ratings. These panels were formed in the areas of Professional Knowledge, Mathematics, Science, Literature/ Fine Arts, Social Studies, Writing, Reading, and Listening. NTE Core Battery scores are not reported at this level of detail, however, so following the determination of the panelists' standards for each test, test scores were aggregated to form scores for each module. Mathematics, Science, Social Studies, and Literature/Fine Arts were combined to form the General Knowledge module. Writing (objective), Writing (essay), Reading, and Listening were included in the Communication Skills module. Scores in this module are weighted so that each test (Listening, Reading, and Writing) have equal weight. Within the Writing Test the objective and essay subtests are weighted equally. Professional Knowledge is a complete module.

At the time that panelists' ratings were aggregated to determine standards, data from the Content Review Panels' ratings were merged with the Knowledge Estimation data. Items judged content appropriate (see Chapter III) were given a weight of 1.0. Content inappropriate items were weighted 0.0. Knowledge Estimation ratings were multiplied by their content appropriateness weight during data processing. Thus, content appropriate items were included in the determination of the standard, but content inappropriate items were dropped from the calculations.

KNOWLEDGE ESTIMATION RESULTS

Essay Ratings

Table 20 presents information concerning the distributions of all essay scores in Louisiana for the November, 1982, administration of the Core Battery and of the sample of papers selected for review. As can be seen in Table 20, a slight oversampling was required in the tails of the distribution (the high and low ends) to provide representation at all score levels. At least two essays were sampled at each score point. Table 21 includes the average number of essays placed in each of the three rating catgories by Content Review judges, Knowledge Estimation judges, and the total group. It can be seen from the Table that there was general agreement between the two types of panels in the percent of essays assigned to each category. Content Review panelists rated somewhat more papers as "Borderline," whereas Knowledge Estimation judges placed more papers in the "Acceptable" category.

Information concerning the average scores assigned to borderline papers is presented in Table 22. Both the mean and median are reported in Table 22. The median is included, since it is a more stable statistic than the mean. In the case in which one judge provides ratings that vary widely from those of other judges, the mean will be more affected by such rating than the median. Inspection of Table 22 indicates that this was not the case. The median was 6.93, and the mean was 6.91. Therefore, the mean was used in all further calculations since it incorporates more information than the median. Based on the results shown in Table 21, it was decided to use results from the total panel in the calculation of the estimated score for the minimally knowledgeable teacher candidate. Note that Tables 21 and 22

TABLE 20

PROPORTION OF LOUISIANA ESSAYS RECEIVING EACH POSSIBLE SCORE NTE CORE BATTERY ADMINISTRATION, NOVEMBER, 1982

Raw Score	Proportion of Louisiana Examinees	Proportion of Sample
12	0.003	0.033
11	0.014	0.033
10	0.058	0.067
9	0.118	0.100
8	0.224	0.200
7	0.196	0.183
6	0.183	0.167
5	0.094	0.083
4	0.069	0.067
3	0.028	0.033
2	0.012	0.033

TABLE 21

AVERAGE PERCENT OF ESSAYS CLASSIFIED IN EACH JUDGMENT CATEGORY BY NTE WRITING PANEL MEMBERS

Panel	Acceptable	Unacceptable	Borderline
Content Review	42.50	24.44	33.06
Knowledge Estimation	48.89	25.28	25.83
Total Panel	44.63	24.72	30.65

present the data by the judges' original task assignments (Content Review or Knowledge Estimation). Half-panel results for Knowledge Estimation are presented later in this chapter.

TABLE 22

MEDIAN AND MEAN OF SCORES ESTABLISHED BY JUDGES AS REPRESENTING BORDERLINE ESSAYS

Panel	Median of Judges' Ratings	Mean of Judges' Ratings
Content Review	6.93	7.04
Knowledge Estimation	6.65	6.65
Total Panel	6.93	6.91

Ebel Results

Results of the Knowledge Estimation Panel ratings of objective test items (the Ebel ratings) are presented in this section. Three tables are provided. Table 23A includes the distribution of relevance ratings by test. The average percent and number of item judgments classified as measuring "essential," "important," or "not very important" knowledge is reported for each form of each test. It can be seen from Table 23A that judges tended to rate items as measuring either "essential" or "not very important" knowledge more often than they rated items as "important." Additionally, judgments were more likely to be in the "not very important" category than any other. While these results were not expected, they are not altogether surprising. Judges were instructed to classify the knowledge required to answer each item corectly as "essential," "important," or "not very important." It is likely that, rather than evaluating the knowledge required to answer an item correctly, judges rated the overall relevance of each item. In isolation, the specific content of a given item (rather than the knowledge it represents) is not likely to be considered essential. In light of the higher than expected percentage of ratings in the "not very important" category, further analyses were conducted. The results appear in Table 23B, which presents the number and percent of items rated Essential or Important by a majority of judges. If over half of the judges for a given test rated an item as either important or essential, that item was included in Table 23B. It is apparent that most of the items in the Core Battery were not considered unimportant using this criterion, in contrast to the percent of judgments tallied in Table 23A. Thus while there was an appreciable number of "not very important" judgments, they tended to be spread across items; very few items received this rating by a majority of judges.

Table 24 follows the same format as Table 23A and reports the difficulty ratings. About one-third of the item judgments across tests were classified as "easy." In six of the eight tests more than 70 percent of the judgments were either "easy" or "moderate" on both Forms A and B. For Form B of the Listening Test, 65.5 percent of the item judgment were "easy" or "moderate," and for Form B of the Literature/Fine Arts Test, 62.0 percent of the ratings fell in these categories.

Calculation of Judges' Standards

There were several steps in the analysis of judges' ratings to calculate an expected minimum score. First, data from the Content Review Panels were merged with the Knowledge Estimation ratings provided by each judge. The ratings of items judged to be content inappropriate were deleted from further analysis at this point. Second, each judge's rating of each item (easy/essential, moderate/essential, etc.) was replaced by the percent of items in the category that the judge predicted would be answered correctly by the minimally knowledgeable teacher candidate. This percentage was then converted to a proportion indicating the probability that the minimally

TABLE 23A

MEAN NUMBER AND PROPORTION OF ITEM JUDGMENTS PLACED IN EACH RELEVANCE CATEGORY BY KNOWLEDGE ESTIMATION PANEL MEMBERS

Test and	Essential		Important		Not Very Important	
Form	Number	Percent[a]	Number	Percent	Number	Percent
Professional Knowledge						
Form A	30.7	29.5	17.4	16.8	56.5	54.3
Form B	32.6	32.0	16.3	15.9	56.1	55.0
Mathematics						
Form A	10.8	43.2	3.2	12.8	10.9	43.6
Form B	11.5	46.0	2.7	10.8	10.7	42.8
Science						
Form A	8.5	28.2	6.9	23.0	14.5	48.2
Form B	7.3	24.2	6.8	22.7	15.8	52.7
Social Studies						
Form A	8.0	26.7	7.4	24.7	14.6	48.6
Form B	10.1	33.6	6.3	21.1	13.6	45.3
Literature and Fine Arts						
Form A	10.2	29.0	6.8	19.3	18.0	51.4
Form B	9.1	26.7	9.1	26.7	16.8	49.3
Reading						
Form A	11.0	36.7	2.7	8.9	16.3	54.4
Form B	11.8	39.2	2.3	7.5	15.9	53.1
Listening						
Form A	18.8	46.9	7.9	19.7	13.1	32.8
Form B	20.0	50.0	5.4	13.6	8.2	20.6
Writing (Objective)						
Form A	15.4	34.1	6.5	14.5	22.9	50.9
Form B	15.4	34.1	5.2	11.5	23.7	52.7

---

[a] Percentages may not total 100 because of rounding and nonresponse.

TABLE 23B

NUMBER AND PERCENT OF ITEMS RATED ESSENTIAL OR IMPORTANT BY A MAJORITY OF JUDGES FOR EACH TEST FORM AND THE AVERAGE PERCENT OF THE TWO TEST FORMS

	FORM A		FORM B		AVERAGE
	Number of Items	Percent of Items	Number of Items	Percent of Items	Percent of Items
Professional Knowledge Test/Module	99	95.2	97	95.1	95.1
Mathematics Test	24	96.0	24	96.0	96.0
Science Test	26	86.7	28	93.3	90.0
Social Studies Test	26	86.7	27	90.0	88.3
Literature/Fine Arts Test	31	88.6	28	82.4	85.5
General Knowledge Module	107	89.2	107	89.9	89.5
Reading Test	30	100.0	30	100.0	100.0
Listening Test	39	97.5	40	100.0	98.7
Writing Test (Objective)	44	97.8	44	97.8	97.9
Communication Skills Module	113	98.3	114	99.1	98.7

knowledgeable teacher candidate would answer that item correctly. As an example, suppose an item had been judged to be moderately difficult and important, and that the judge predicted that 50 percent of such items would be answered correctly by the minimally knowledgeable teacher candidate. The probability of a minimally knowledgeable teacher candidate answering an item in this category correctly would be .50. The sum of such probabilities for all items in the test is the predicted raw score (the number of items answered correctly) of that judge. The predicted raw scores were averaged across judges for each test. These average test scores were finally transformed to scale scores for the three modules of the Core Battery using transformation formulas provided by Educational Testing Service. The resulting scaled score standards are presented in Table 25.

TABLE 24

MEAN NUMBER AND PROPORTION OF ITEM JUDGMENTS PLACED IN EACH DIFFICULTY CATEGORY BY KNOWLEDGE ESTIMATION PANEL MEMBERS

Test and	Easy		Moderate		Hard	
Form	Number	Percent[a]	Number	Percent	Number	Percent
Professional Knowledge						
Form A	27.5	26.4	48.4	46.5	28.1	27.0
Form B	28.9	28.4	45.4	44.5	27.6	27.0
Mathematics						
Form A	9.1	36.4	10.0	40.0	5.8	23.2
Form B	9.0	36.0	10.6	42.4	5.3	21.2
Science						
Form A	11.5	38.5	12.3	40.9	6.0	20.0
Form B	8.8	29.4	13.2	43.9	7.9	26.4
Social Studies						
Form A	12.7	42.2	10.6	35.3	6.8	22.5
Form B	13.9	46.4	10.7	35.6	5.4	18.1
Literature and Fine Arts						
Form A	11.8	33.6	14.6	41.7	8.6	24.5
Form B	8.5	25.0	12.6	37.0	12.9	38.0
Reading						
Form A	9.4	31.4	14.3	47.8	6.3	20.8
Form B	9.3	31.1	15.6	51.9	5.0	16.7
Listening						
Form A	12.9	32.2	18.6	46.4	8.4	21.1
Form B	11.4	28.6	14.8	36.9	6.2	15.6
Writing (Objective)						
Form A	10.9	24.2	20.7	46.1	13.3	29.5
Form B	12.9	28.7	21.1	46.9	10.5	23.4

---

[a] Percentages may not total 100 because of rounding and nonresponse.

The results for Form A and Form B are similar for the Professional Knowledge and General Knowledge modules, with a range of no more than three scaled points between Form A and Form B. However, the range between Form A and Form B of the Communication Skills module was greater. This was largely due to differences in the number of judges rating the two forms of the Listening Test. One judge rated most of the items on both forms as "not very important." The impact of this judge's ratings was greater on Form B than Form A because fewer judges evaluated Form B, and resulted in a lower standard for the total Form B Communication Skills Module.

CONSISTENCY OF RESULTS FOR KNOWLEDGE ESTIMATION PROCESS

The half-panel analyses for Knowledge Estimation changed from the method employed in the 1978 validation study to match the changes in the procedures used for producing Knowledge Estimation ratings. The half-panel analyses presented here consisted of four parts: a) a comparison of the half-panels' determination of Ebel methodology standards by module and test

TABLE 25

RESULTS OF APPLICATION OF EBEL METHODOLOGY
STANDARDS BY NTE CORE BATTERY MODULE

Module	Maximum Score Possible	Form A	Form B	Average
Professional Knowledge	690	654	651	652
General Knowledge	690	652	650	651
Communication Skills	690	656	647	652

form; b) a comparison of half-panel ratings of item difficulty by test and test form; c) a comparison of half-panel ratings of item relevance by test and test form; and d) a comparison of half-panel ratings of the percent of items the minimally knowledgeable teacher candidate would be expected to answer correctly by test.

In order to apply the Ebel methodology, the judges gave ratings for item difficulty, item relevance, and the percentage of items in each combination of difficulty and relevance that the minimally knowledgeable teacher candidate would answer correctly. The half-panel analyses began with the end-product of these judgments and worked back to contrast each component part across half-panels. This procedure served two functions: it presented the consistency of results for the overall process and it examined the sources of inconsistencies when they appeared.

Consistency of Ebel Results

The final results from the derivation of the Ebel methodology standards by half-panel are shown in Table 26. Half-panel estimates of the standard scaled score that could be expected of the minimally knowledgeable teacher candidate are shown for both Form A and Form B of each module.

Differences in half-panel estimates ranged from nine points (Form A of the Communication Skill module) to three points (both forms of General Knowledge). T-tests of the differences between raw scores assigned by the two half-panels to the eight tests comprising these modules indicated that none of the differences between half-panel ratings were statistically significant.

TABLE 26

RESULTS OF APPLICATION OF EBEL METHODOLOGY STANDARDS BY NTE CORE BATTERY MODULE, FORM, AND HALF-PANEL: SCALE SCORES

	Form A		Form B		Both Forms	
	Panel 1	Panel 2	Panel 1	Panel 2	Panel 1	Panel 2
Professional Knowledge	651	656	649	654	650	655
General Knowledge	650	653	648	651	649	652
Communication Skills	652	661	651	658	652	660

After determining that the overall results from the Ebel methodology standards were consistent across half-panels, each task in the procedure was then examined to determine if this consistency had been maintained throughout the standard setting process. Two analyses of the item difficulty ratings are shown by half-panel in Table 27. These are Pearson product-moment correlations of the half-panels' judgments of item difficulty and average differences in the half-panels' ratings of item difficulty.

The Pearson product-moment correlation values between half-panel judgments of item difficulty ranged from .39 (Form A of the Listening objective subtest) to .73 (Form B of the Mathematics Test). All correlations were significant at the $p<.01$ level or greater, and 13 of the 16 correlations were significant at the level of $p<.001$. It was concluded from this test that the half-panel ratings for each of the Core Battery tests showed a high degree of association.

The differences in half-panel ratings of item difficulty could range from 0.00 (if the average difficulty ratings were identical for both half-

TABLE 27

COMPARISON OF HALF-PANEL RATINGS ON ITEM DIFFICULTY BY TEST FORM

Examinations	Form A		Form B	
	Correlation Between Half-Panel Judgments on Items[a]	Mean Difference in Half-Panel Scores on Items[b]	Correlation Between Half-Panel Judgments on Items[a]	Mean Difference in Half-Panel Scores on Items[b]
Professional Knowledge	.67	.04	.65	.09
Mathematics	.69	.11	.73	.24
Science	.68	.11	.62	.01
Social Studies	.63	.15	.53	.01
Literature/Fine Arts	.69	.19	.55	.21
Reading	.57	.13	.48	.15
Writing (Objective)	.72	.33	.57	.31
Listening[c]	.39	.13	.46	.20

[a] Pearson product-moment correlations are reported. All correlations were significant at the $p < .05$ level, except for the correlations for Forms A and B for the Listening examination. These two correlations (-.02, .08) were not significant.

[b] The differences reported here were computed by comparing the average difficulty rating for items given by Panel 1 and Panel 2. The range for these differences was 0 to 2, since the minimum relevance rating was 1.0 and the maximum revelance rating was 3.0.

[c] Note that half-panel composition is different for Forms A and B for Listening as explained in Chapter II.

panels) to 2.00 (if one half-panel gave an average difficulty rating of 1.00, or "easy," and the other half-panel gave an average difficulty rating of 3.00, or "hard"). The range actually observed for these differences was from .01 (Form B of the Science and Social Studies Tests) to .33 (Form A of the Writing Objective subtest). The average difference between half-panel ratings across all tests was .15.

Similar comparisons of consistency for item relevance ratings are shown in Table 28. The lowest correlations reported are for the Listening Test: .02 for Form A and .08 for Form B. The remaining correlations ranged from .35 (Form B of Literature/Fine Arts) to .85 (Form B of Mathematics). All correlations, with the exception of those for the Listening Test, were significant at the $p<.05$ or greater level. Eight of the correlations were significant at the level of $p<.001$, indicating a high degree of association between half-panel ratings of item relevance.

The overall correlation values for item relevance are very similar to those for item difficulty, with the exception of the item relevance correlations for the Listening Test. These low correlations, noted in the preceding paragraph, occurred because one panelist rated most of the items in the Listening Test as "not very important." The effect of this panelist's ratings was more pronounced on Form B, for which there were seven judges, than on Form A, for which there were nine judges. There were no procedural reasons for excluding this panelist's judgments, but the half-panel correlations for the Listening Test should be interpreted with the understanding that they include one set of opinions that diverged widely from those of the other judges.

TABLE 28

COMPARISON OF HALF-PANEL RATINGS ON ITEM RELEVANCE BY TEST FORM

Examinations	Form A		Form B	
	Correlation Between Half Panels' Judgments on Items[a]	Mean Difference in Half Panels' Scores on Items[b]	Correlation Between Half Panels' Judgments on Items[a]	Mean Difference in Half Panels' Scores on Items[b]
Professional Knowledge	.59	.22	.53	.23
Mathematics	.61	.27	.85	.35
Science	.59	.09	.43	.11
Social Studies	.50	.23	.44	.15
Literature/Fine Arts	.68	.08	.35	.04
Reading	.47	.03	.37	.06
Writing (Objective)	.53	.14	.56	.24
Listening[c]	-.02	.42	.08	.61

[a] Pearson product-moment correlations are reported. All correlations were significant at the $p < .05$ level, except for the correlations for Forms A and B for the Listening examination. These two correlations (-.02, .08) were not significant.

[b] The differences reported here were computed by comparing the average relevance rating for items given by Panel 1 and Panel 2. The range for these differences was 0 to 2, since the minimum relevance rating was 1.0 and the maximum revelance rating was 3.0.

[c] Note that half-panel composition is different for Forms A and B for Listening as explained in Chapter II.

The difference between half-panel ratings of item relevance could range from 0.00 to 2.00, as they could for item difficulty. A difference of 0.00 would mean that the two half-panels made identical average ratings for item relevance. A difference of 2.00 would mean that one half-panel judged all items to be "not very important" (1.00), and the other half-panel judged all items to be "essential" (3.00). The observed range for these differences was highest for the Listening Test: .42 for Form A and .61 for Form B. Among the other tests the range was from .03 (for Form A of the Reading Test) to .35 (for Form B of the Mathematics Test). Excluding the Listening Test, the average difference across the remaining seven tests was .15. This was the same average difference as that found for item difficulty.

The final comparison was that between half-panel estimates of the percent of items the minimally knowledgeable teacher candidate would answer correctly. That information is presented in Table 29. The values for the half-panel correlations were consistently high, ranging from .88 on the Listening Test to .97 on the Professional Knowledge and Literature/Fine Arts Tests. For half-panel differences in percentage estimates, the values could range from 0.00 (complete average agreement) to 1.00 (complete disagreement, on the average, for all nine categories of difficulty and relevance). The observed range was from .02 for the Literature/Fine Arts Test to .20 for the Mathematics Test. The average difference across tests was .10.

Consistency of Essay Results

The consistency measures described above were not appropriate for the essay subtest of the Writing Test, because the Knowledge Estimation procedures used here differed from those employed with the other tests. Consis-

TABLE 29

COMPARISON OF HALF-PANEL RATINGS ON PERCENT OF ITEMS THE MINIMALLY KNOWLEDGEABLE TEACHER CANDIDATE WOULD BE EXPECTED TO ANSWER CORRECTLY

	Correlation Between Half-Panels' Judgments Across Categories[a]	Difference Between Half-Panels' Scores Across Categories[b]
Professional Knowledge	.97	.07
Mathematics	.95	.20
Science	.94	.13
Social Studies	.96	.10
Literature/Fine Arts	.97	.02
Reading	.90	.13
Writing (Objective)	.90	.04
Listening	.88	.14

[a] The correlations reported here are Pearson product-moment correlations. All correlations were significant at the $p < .01$ level.

[b] These differences were computed by comparing the average percentage difference in ratings given by Panel 1 and Panel 2 for the nine categories of item difficulty/relevance. The range for these differences was 0 to 1, since the maximum difference would be between 0 percent (.00) and 100 percent (1.00).

TABLE 30

AVERAGE PERCENT OF ESSAYS CLASSIFIED IN EACH JUDGMENT CATEGORY BY HALF-PANEL

Panel	Mean Score	Percent Acceptable	Percent Unacceptable	Percent Borderline
Panel 1	6.76	47.59	20.74	31.67
Panel 2	7.06	41.67	28.70	29.63
Total Panels	6.91	44.63	24.72	30.65

tency was measured here by comparing the number of panelists in each half-panel who placed essays in the three categories of "acceptable," "unacceptable," or "borderline." The distributions shown in Table 30 are similar for the two half-panels. Members of Panel 2 placed approximately 8 percent more essays in the "unacceptable" category, while those in Panel 1 rated approximately 6 percent more essays as "acceptable." The average essay scores produced by the two half-panels ranged from 6.76 (Panel 1) to 7.06 (Panel 2), a difference of .30 points on a range of 0 to 12. A T-test of this difference found it to be statistically nonsignificant.

Conclusions

The data presented in measuring Knowledge Estimation Panel consistency are interpreted as showing a high degree of agreement between half-panels about item difficulty, item relevance, and the score that could be expected of the minimally knowledgeable teacher candidate. The one exception to this general agreement occurred in the relevance ratings for the Listening Test.

REFERENCES CITED

Angoff, W. H. Scales, norms, and grade equivalent scores. In R. L. Thorndike (Ed.). *Educational measurement* (2nd Edition). Washington D.C.: American Council on Education, 1971.

Ebel, R. L. *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice Hall, Inc., 1972.

Educational Testing Service. *Report on a Study of the National Teacher Examinations by the State of Louisiana*. Princeton, New Jersey: Educational Testing Service, 1978.

Hambleton, R. K. & Eignor, D. R. A practitioner's guide to the development and validation of criterion-referenced tests. *Laboratory of Psychometric and Evaluative Research Report No. 70*. Amherst, MA: School of Education, University of Massachusetts, 1979.

Hambleton, R. K. & Eignor, D. R. Competency test development, validation, and standard setting. In R. M. Jaeger & C. K. Tittle (Eds.). *Minimum competency achievement testing: Motives, models, measures and consequences*. Berkeley, CA: McCutchan, 1980.

Jaeger, R. M. Measurement consequences of selected standard-setting models. *Florida Journal of Educational Research*, 1976, *18*, 22-27.

Livingston, S. A. & Zieky, M. J. *Passing scores*. Princeton, NJ: Educational Testing Service, 1982.

Nedelsky, L. Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 1954, *14*, 3-19.

Popham, W. J. *Criterion-referenced measurement*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

Shepard, L. A. Setting standards and living with them. *Florida Journal of Educational Research*, 1976, *18*, 23-32.

CHAPTER V

INFORMATION ABOUT RISK AND DEMAND

The data presented to this point have addressed two fundamental considerations in establishing a rational basis for the use of NTE Core Battery scores in the certification process. These are: a) the relationship between the content of the tests and the content of the teacher training curricula in Louisiana, and b) the estimated score level that would be achieved by a minimally knowledgeable teacher candidate.

Information will now be presented concerning two other aspects of applying NTE scores in the certification process. First, when the score standards derived from the Validation Team's judgments are adjusted to allow for certification of the maximal number of qualified applicants, the risk of certifying unqualified applicants is inevitably affected. The level of risk (i.e., rejecting qualified applicants or accepting unqualified applicants) associated with each adjustment is included in this chapter. Second, the standard eventually adopted for certification will affect the supply of new teachers in Louisiana. For the present study, Educational Testing Service has supplied the performance distributions of Louisiana examinees taking the Core Battery in November, 1982, and March, 1983. To the extent that these examinees are typical of all applicants for initial certification in the State, these data can be used to estimate the probable effect that use of a given set of score standards will have on the supply of teachers applying for certification. Before addressing these two additional concerns, however, the format of the revised NTE qualifying score will be discussed.

FORMAT OF REVISED NTE QUALIFYING SCORE

When the Common Examinations were used for teacher certification purposes in Louisiana from 1978-1982, the format of the qualifying score was the combined Area Examination and Weighted Common Examinations Total (WCET) scores. In other words, a candidate's Area and WCET scores were summed, and the total was compared to a composite standard. This model had two salient features: compensation (a high WCET score could compensate for a low Area score, and vice versa), and an approximately equal weighting of the Area and WCET scores. Since the WCET was reported as a single score, the Area score carried greater weight than any single component subtest of the Common Examinations. Within the WCET, the following weights were assigned by Educational Testing Service to each of the component subtests: Professional Education (4.0), Written English Expression (1.0), Social Studies and Literature/Fine Arts (2.5), and Science and Mathematics (2.5).

As mentioned previously, the Core Battery has a format different from that of the Common Examinations. The Core Battery consists of three two-hour modules, including a total of eight tests; the Commons was a single three and one-quarter hour examination comprised of six tests. Educational Testing Service will report each of the three Core Battery module scores separately and will not weight the modules differentially or combine them into a single Core Battery composite score. Furthermore, the scaling of each module differs from the Area and Common Examinations scales: the range of scaled scores for the Core Battery modules is 600-690, while the Common and Area Examinations' score ranges were 250-990.

The NTE Blue Ribbon Score Committee met on March 22, 1983, to recommend a format for the Core Battery and Area Examinations standards. Although the Area Examinations were not revised (and therefore were not the subject of the present validation study), the application of Area scores in the certification process could be affected by the differences between the Common Examinations and Core Battery previously noted. The Blue Ribbon Score Committee considered a number of format options for the new qualifying score(s), including single versus multiple, weighted versus unweighted, and compensating versus noncompensating score models. After a lengthy discussion and the consideration of several alternatives, the committee voted in favor of a four-score, noncompensating format for the revised NTE standards. According to this model, each Core Battery module and the appropriate Area Examination would have an independent minimum score that a teacher candidate must meet or surpass in order to be eligible for certification. The remainder of the discussion of qualifying scores will reflect this structural change from the former NTE composite score.

RISK OF REJECTING QUALIFIED CANDIDATES

Each test score identified by the judges as the level that a minimally knowledgeable teacher candidate would achieve is a theoretical true score. The score actually achieved by any person taking the NTE falls within a range of scores on either side of the examinee's true score. One convenient way to interpret the score is to recall that, under certain assumptions, there are 68 chances in 100 that the examinee's observed score will be within one standard error of measurement [1] of his or her true score, 95 chances in 100 that the observed score will be within two standard errors of measurement of his or her true score, and 99 chances in 100 that it will be within three standard errors of measurement of his or her true score. If the scores identified by the Validation Team judges are adopted as the standards for the Core Battery modules, some applicants with test scores below the standards will be rejected even though their true scores, if known, would be above the standard. The reverse is also true. Thus, decisions in establishing the NTE standard should take into account the probability at different score levels of rejecting a truly qualified applicant.

Table 31 presents the probabilities associated with the risk of rejecting qualified candidates with true scores greater than a specified value, and of accepting unqualified candidates with true scores of less than that value. The probabilities are given for several standards of acceptance expressed in multiples of the standard error of measurement (SEM). The particular multiples of the standard error of measurement in Table 31 are merely illustrative; any multiple can be used. The values in this table are not specific to any particular test, but are generalized probabilities.

An example may help to illustrate how Table 31 can be used. Suppose one were to consider adjusting the study estimate by subtracting one SEM from it. In this case one would locate the SEM value of -1

---

[1] The standard error of measurement is an estimate of the amount of variation in a performance measure (e.g., test score) attributable to measurement error. It is theoretically equal to the mean difference between examinees' observed score and true scores.

TABLE 31

PROBABILITES OF INCORRECTLY REJECTING QUALIFIED APPLICANTS AND INCORRECTLY ACCEPTING UNQUALIFIED APPLICANTS AT SIX LEVELS OF THE STANDARD ERROR OF MEASUREMENT (SEM)

	Probability of Rejecting Applicant Whose True Score is the Specified Number of SEMs above Standard						Probability of Accepting Applicant Whose True Score is the Specified Number of SEMs below Standard					
SEM	0.0	0.5	1.0	1.5	2.0	2.5	0.0	0.5	1.0	1.5	2.0	2.5
0.0	.50	.31	.16	.07	.02	.01	.50	.31	.16	.07	.02	.01
-0.5	.31	.16	.07	.02	.01	.001	.69	.50	.31	.16	.07	.02
-1.0	.16	.07	.02	.01	.001	<.001	.84	.69	.50	.31	.16	.07
-1.5	.07	.02	.01	.001	<.001	<.001	.93	.84	.69	.50	.31	.16
-2.0	.02	.01	.001	<.001	<.001	<.001	.98	.93	.84	.69	.50	.31
-2.5	.01	.001	<.001	<.001	<.001	<.001	.99	.98	.93	.84	.69	.50

(third row in the Table). The left side of the Table indicates that there would then be 16 chances in 100 of failing to certify a candidate whose true score was equal to the specified score, seven chances in 100 of failing to certify a candidate whose true score was one-half of a SEM above the specified score, and so on. The right side of the Table indicates that there would be 84 chances in 100 of certifying a candidate whose true score was equal to the specified score, 69 chances in 100 of certifying an applicant whose true score was one-half of a SEM below the specified score, and so on.

Table 32 presents the mean score and standard deviation for each of the Core Battery modules for those persons who attempted the NTE in the fall of 1982 or the spring of 1983 and who identified themselves as seniors at the time of testing. The data are also limited to examinees who took the entire NTE (Three Core Battery modules and an appropriate

Area Examination). Included in Table 32 are the percent of examinees achieving or exceeding the minimum score proposed by the Validation Team judges on each module and the standard error of measurement (SEM) for each module. This Table is limited to the percent of examinees meeting each individual standard. The joint passing rates for all four standards as a whole are presented in subsequent tables.

TABLE 32

MEAN EXAMINEE SCORE AND PERCENT OF EXAMINEES MEETING JUDGES' STANDARDS ON NTE CORE BATTERY MODULES

	Judges' Minimum Score	SEM	Mean Examinee Score (n=1013)	Standard Deviation	Percent at or above Minimum
Professional Knowledge	652	3.8	656.5	11.3	67.4
General Knowledge	651	3.5	655.7	12.3	65.6
Communication Skills	652	3.5	658.9	11.4	74.8

NOTE: SEM derived from fall, 1982, form of Core Battery.

Table 33 gives the percentages of Louisiana examinees scoring above the study estimate of the score standards for the three Core Battery modules and each Area Examination. Tables 34 to 36 give the percentages of Louisiana examinees scoring above the study estimate of the score standards for the three Core Battery modules minus one, two, and three standard errors of measurement, respectively.

TABLE 33

PERCENTAGE OF LOUISIANA EXAMINEES SCORING AT OR ABOVE STUDY ESTIMATE OF CORE BATTERY SCORE STANDARDS BY AREA EXAMINATION

Area Examination	Number of Examinees	Standard for Area	Estimate for PK	Estimate for GK	Estimate for CS	% Pass
Agriculture	6	466	652	651	652	33.3
Biology & General Science	23	575	652	651	652	69.6
Business Education	40	591	652	651	652	45.0
Chemistry/Physics/ General Science	3	530	652	651	652	100.0
Early Childhood Education	96	506	652	651	652	43.8
Education in Elementary Schl	511	545	652	651	652	54.0
Education of Mental Retarded	12	541	652	651	652	33.3
English Lang/ Literature	49	441	652	651	652	83.7
French	2	517	652	651	652	50.0
German	0	496	652	651	652	--
Home Economics Education	29	509	652	651	652	55.2
Mathematics	25	617	652	651	652	28.0
Media Specialist/ Library/A-V	0	563	652	651	652	--
Music Education	46	533	652	651	652	58.7
Physical Education	113	545	652	651	652	36.3
Social Studies	46	563	652	651	652	52.2
Spanish	4	538	652	651	652	75.0
Speech Communic & Theatre (OLD)	8	519	652	651	652	50.0
Total Across Areas	1013	----	652	651	652	51.8

PK=Professional Knowledge; GK=General Knowledge; CS=Communication Skills

NOTE: See text for explanation of Area Examination Standards.

TABLE 34

PERCENTAGE OF LOUISIANA EXAMINEES SCORING AT OR ABOVE STUDY ESTIMATE OF CORE BATTERY SCORE STANDARDS MINUS ONE SEM BY AREA EXAMINATION

Area Examination	Number of Examinees	Standard for Area	Estimate for PK	Estimate for GK	Estimate for CS	% Pass
Agriculture	6	466	649	648	649	50.0
Biology & General Science	23	575	649	648	649	73.9
Business Education	40	591	649	648	649	55.0
Chemistry/Physics/ General Science	3	530	649	648	649	100.0
Early Childhood Education	96	506	649	648	649	50.0
Education in Elementary Schl	511	545	649	648	649	62.4
Education of Mental Retarded	12	541	649	648	649	50.0
English Lang/ Literature	49	441	649	648	649	91.8
French	2	517	649	648	649	50.0
German	0	496	649	648	649	--
Home Economics Education	29	509	649	648	649	65.5
Mathematics	25	617	649	648	649	32.0
Media Specialist/ Library/A-V	0	563	649	648	649	--
Music Education	46	533	649	648	649	69.6
Physical Education	113	545	649	648	649	47.8
Social Studies	46	563	649	648	649	54.3
Spanish	4	538	649	648	649	75.0
Speech Communic & Theatre (OLD)	8	519	649	648	649	75.0
Total Across Areas	1013	----	649	648	649	60.3

PK=Professional Knowledge; GK=General Knowledge; CS=Communication Skills

NOTE: See text for explanation of Area Examination Standards.

TABLE 35

PERCENTAGE OF LOUISIANA EXAMINEES SCORING AT OR ABOVE STUDY ESTIMATE OF CORE BATTERY SCORE STANDARDS MINUS TWO SEMS BY AREA EXAMINATION

Area Examination	Number of Examinees	Standard for Area	Estimate for PK	Estimate for GK	Estimate for CS	% Pass
Agriculture	6	466	645	644	645	83.3
Biology & General Science	23	575	645	644	645	78.3
Business Education	40	591	645	644	645	57.5
Chemistry/Physics/ General Science	3	530	645	644	645	100.0
Early Childhood Education	96	506	645	644	645	63.5
Education in Elementary Schl	511	545	645	644	645	68.7
Education of Mental Retarded	12	541	645	644	645	50.0
English Lang/ Literature	49	441	645	644	645	91.8
French	2	517	645	644	645	50.0
German	0	496	645	644	645	--
Home Economics Education	29	509	645	644	645	86.2
Mathematics	25	617	645	644	645	32.0
Media Specialist/ Library/A-V	0	563	645	644	645	--
Music Education	46	533	645	644	645	80.4
Physical Education	113	545	645	644	645	66.4
Social Studies	46	563	645	644	645	58.7
Spanish	4	538	645	644	645	75.0
Speech Communic & Theatre (OLD)	8	519	645	644	645	75.0
Total Across Areas	1013	----	645	644	645	68.5

PK=Professional Knowledge; GK=General Knowledge; CS=Communication Skills

NOTE: See text for explanation of Area Examination Standards.

TABLE 36

PERCENTAGE OF LOUISIANA EXAMINEES SCORING AT OR ABOVE STUDY ESTIMATE OF CORE BATTERY SCORE STANDARDS MINUS THREE SEMS BY AREA EXAMINATION

Area Examination	Number of Examinees	Standard for Area	Estimate for PK	Estimate for GK	Estimate for CS	% Pass
Agriculture	6	466	641	641	642	83.3
Biology & General Science	23	575	641	641	642	78.3
Business Education	40	591	641	641	642	60.0
Chemistry/Physics/ General Science	3	530	641	641	642	100.0
Early Childhood Education	96	506	641	641	642	72.9
Education in Elementary Schl	511	545	641	641	642	73.2
Education of Mental Retarded	12	541	641	641	642	58.3
English Lang/ Literature	49	441	641	641	642	93.9
French	2	517	641	641	642	50.0
German	0	496	641	641	642	--
Home Economics Education	29	509	641	641	642	93.1
Mathematics	25	617	641	641	642	32.0
Media Specialist/ Library/A-V	0	563	641	641	642	--
Music Education	46	533	641	641	642	82.6
Physical Education	113	545	641	641	642	75.2
Social Studies	46	563	641	641	642	58.7
Spanish	4	538	641	641	642	75.0
Speech Communic & Theatre (OLD)	8	519	641	641	642	75.0
Total Across Areas	1013	----	641	641	642	73.2

PK=Professional Knowledge; GK=General Knowledge; CS=Communication Skills

NOTE: See text for explanation of Area Examination Standards.

The standards for the Area Examinations in all four tables were derived from the 1978 validation study estimates. These estimates were not used to establish independent standards for the Area Examinations in the 1978 composite score model. In order to allow for the effect of the transition from the former composite (conpensating) model to the present noncomposite model, a statistical adjustment of three SEMs from the 1978 study estimates was necessary. These Area Examination standards appear as constants in Tables 33 through 36 and the remainder of this report.

Only those Area Examinations that are required for certification in Louisiana and that were judged to be valid in the 1978 study are included in the analyses presented in this chapter. Candidates seeking teacher certification in the areas of Speech Pathology, Art Education and Industrial Arts Education will be required to meet the standards established on the three Core Battery modules but will not be required to meet an Area Examination standard. Candidates graduating from the Generic Special Education programs in Louisiana will also be required to pass only the three Core Battery modules.

Educational Testing Service has also revised the Speech Communications and Theatre test, producing a new Speech Communications Area Examination to replace it. Candidates seeking certification in this area and taking the new Speech Communications Area Examination will be required to meet only the Core Battery module standards until the new Speech Communications Area Examination has been studied for validity and a standard set for the test. Candidates seeking certification in Speech Communications after the new standard has been set will be required to meet the Area Examination and Core Battery module standards. A validation study of the new Speech Communications Area Examination is scheduled for the fall of 1983.

It should also be noted on Tables 33 through 36 that no Louisiana examinees attempted the German or the Media Specialist-Library Audio and Visual Services Area Examinations in the fall of 1982 or the spring of 1983. As a result Tables 33 through 36 do not report estimates of the percent of examinees expected to meet the proposed Core Battery module standards in these areas.

SUPPLY OF NEW TEACHERS IN LOUISIANA

Closely related to the consideration of the risk of rejecting qualified candidates is the assessment of the effect of proposed qualifying score requirements on the supply of new teachers in each of the specialty fields. If there is a substantial shortage of teachers at a particular time, a higher risk of certifying unqualified candidates may be considered acceptable. Information on teacher supply and demand may be used in conjunction with the validation data reported above to estimate the effect of any set of score standards on the probable supply of candidates and, indirectly, on the capability of meeting demand.

Tables 33 through 36 give selected statistics based on Louisiana seniors who took the NTE Core Battery in the fall of 1982 or the spring of 1983. If an examinee had taken any of the Core Battery modules or the Area Examination more than once, only the first scores earned were included. This procedure is consistent with the 1978 NTE validation study in Louisiana. There was one exception. If an examinee who had previously taken an Area Examination attempted it again in April, 1983, the more recent Area Examination Score was used.

The values of the means of the estimated scores cannot be compared across teaching areas, since the scores on different Area Examinations are not on comparable scales. The percentages of examinees whose scores exceed the estimated means can be compared, although in interpreting the comparisons it is essential to recognize that different fields may attract students with different levels of talent. Moreover, the numbers of examinees who take some of the Area Examinations are relatively small, so that the results for a different period could vary appreciably from those observed.

As shown in Table 33, the application, without adjustment of the NTE score levels suggested by this study, would result in the certification of different numbers and percentages of candidates in different teaching fields. On the average, only 51.8 percent of the 1013 examinees whose scores were tabulated would be certified accordingly. Historically, validation study estimates of performance standards often yield actual passing rates that are unrealistically low, given the demand for new teachers in many states. In keeping with the report of the 1978 study and its implementation by the Blue Ribbon Score Committee and Superintendent of Education at that time, only deviations from the study estimate that yield higher passing rates (i.e., by subtracting standard errors of measurement) are included in the present report. The effect on passing rates of standards that are more stringent than the study estimates (i.e., by adding standard errors of measurement) will be available to the Blue Ribbon Score Committee and the Superintendent of Education, if and when requested.

Finally, it should be restated that the projected passing rates presented in this chapter are based upon the 1013 Louisiana seniors who were examined on the Core Battery modules and an appropriate Area Examination. This method of analysis is consistent with the 1978 validation study. However, the total number of individuals seeking teacher certification in a given year may exceed the 1013 included in this study's population. In comparing the State's need for teachers with the percent of candidates in the validation population who would meet a given set of standards, it is important to recognize that the total number of certifiable teacher candidates will almost certainly exceed the number of college senior teacher candidates who would pass each set of standards. Thus, while the passing rates discussed in Tables 33 through 36 apply to graduating senior teacher candidates, the actual number of certifiable teachers will include persons in addition to college seniors and will very probably be somewhat higher than the numbers shown here.

CHAPTER VI

SUMMARY OF THE STUDY

This chapter is a summary of the preceding text. It is intended as a brief synopsis of the findings and an overview of the entire study. Earlier chapters should be reviewed for complete documentation and more detailed explanation of the findings summarized below.

STUDY DESIGN

The purpose of this study was to evaluate the validity of the NTE Core Battery for use in certifying teachers in Louisiana. The design was modelled on the one developed by Educational Testing Service and used in Louisiana's 1978 validation study of the entire NTE. In the current design, panels of faculty from Louisiana colleges and universities with teacher education programs reviewed the eight tests comprising the NTE Core Battery. They provided judgments about the following major aspects of the Core Battery as an instrument for teacher certification.

> Content Review. Is the emphasis given to topics within each test the same as the emphasis these topics receive in Louisiana teacher education curricula? Would graduates from Louisiana teacher education programs have had the opportunity to learn the content included in each test item?
>
> Knowledge Estimation. Would the minimally knowledgeable teacher candidate find each test item easy, moderately difficult, or hard? Is the knowledge measured in each test item essential, important, or not very important? What proportion of items in each of the nine resulting categories of difficulty and relevance would the minimally knowledgeable teacher candidate answer correctly?

The task of the Content Review Panel was to provide information regarding the match between the content of the tests and teacher education curricula. This ensured that the estimated scores would include only those items that were appropriate measures of Louisiana teacher education curricula; that is, those that reflect content the teacher candidates would have had the opportunity to learn. The judgments of this Panel were combined with those of the Knowledge Estimation Panel to produce estimates of the score that could be expected of the minimally knowledgeable teacher candidate on each Core Battery test. The resulting scores were then combined to produce estimated scores for the three modules of the Core Battery: Professional Knowledge, General Knowledge, and Communication Skills. The Blue Ribbon Score Committee also met during the time the study was under way and recommended that the required NTE score take the form of four minimum scores (one for each of the three Core Battery modules plus the appropriate Area Examination) in determining a teacher certification standard.

STUDY PARTICIPANTS

The usefulness of the study design rested largely upon the degree to which the faculty who served on the Content Review and Knowledge Estimation Panels was representative of Louisiana teacher education programs. As Chapter II illustrates, this objective was achieved. Twenty-one of the 22 colleges and universities with such programs were included in the Panels, and the number of faculty participating from each institution matched the institution's size in the number of graduating students. The Panels also reflected an equitable representation of private institutions and those with predominantly black student bodies. For all Core Battery tests the number of faculty participating in both the Content Review and Knowledge Estimation Panels met or exceeded the number required for reliable results.

CONTENT REVIEW FINDINGS

The results of the Content Review Panel judgments are presented here for the three Core Battery modules and for each of the eight component tests. These summaries include the content appropriateness of test items, the comprehensiveness of the tests, relative emphasis given topics in the tests and in Louisiana teacher education curricula, and the overall similarity between the tests and these curricula. Chapter III provides a complete discussion of the procedures and data upon which these summaries are based.

Professional Knowledge Module

This module consisted of one test, Professional Knowledge. It was judged to be very closely related to the Louisiana teacher education curricula. Ninety-three percent of the items were judged appropriate, and there were no topics noted by two or more judges as omitted from the test. The emphasis given topics within the test was acceptable. When there was a difference in emphasis, topics tended to be overrepresented on the test in comparison with the emphasis they received in teacher education programs. Eighty-eight percent of the judges felt that the test closely paralleled or did not differ significantly from Louisiana teacher education curricula.

General Knowledge Module

The General Knowledge Module included four tests: Mathematics, Science, Social Studies, and Literature/Fine Arts. The module was evaluated as closely related to Louisiana teacher education curricula. Each component test is discussed below.

Mathematics Test

Ninety-eight percent of the items were judged content appropriate. Three topics (probability, statistics, and algebraic problems) were cited as omitted. The relative emphasis was acceptable, and where there were differences between the test and curricula, topics were overrepresented on the test. Eighty percent of the judges rated the content of the test as a close match with the teacher education curricula.

Science Test

Ninety-seven percent of the items were judged content appropriate. No topics were noted by two or more panelists as omitted. The relative emphasis was acceptable, and where differences were noted, topics were generally overrepresented on the test. The test was perceived as matching the content of Louisiana teacher education programs by 90 percent of the panelists.

Social Studies Test

In this test 100 percent of the items were judged to be content appropriate. No topic was listed by two or more panelists as omitted from the test. The value for relative emphasis was .48, less than the .50 level set for exclusion from additional review on this criterion. Again, when panelists felt that the emphasis given topics on the test differed from that given topics in teacher education curricula, they judged the topics to be overrepresented on the test. However, some 82 percent of the panelists evaluated the test as closely paralleling, or not differing significantly from, Louisiana teacher education programs.

Literature/Fine Arts Test

Eighty-four percent of the items were evaluated as content appropriate. Five of the 10 items judged inappropriate dealt with a single topic, "relating works of art to one another and their social-historical context." One topic, recognition of the names and works of famous people, was noted by two judges as omitted from the test. The relative emphasis was .43, less than the .50 set as an acceptable level. When there was a difference between the emphasis topics received on the test and in Louisiana teacher education curricula, topics were generally overrepresented on the test. Sixty-seven percent of the panelists judged that the content of the test matched that of teacher education curricula closely or with no significant differences.

Communication Skills Module

This module was judged to be closely related to Louisiana teacher education programs. It included three tests: Reading, Listening, and Writing. The Writing Test was, in turn, composed of an objective subtest and an essay exercise. The essay component was not considered by the Content Review Panel because it did not include discrete items that could be evaluated for their appropriateness or for their match with the emphasis topics received in teacher education curricula.

Reading Test

All of the items (100%) were evaluated as content appropriate, and no topic was felt to be omitted. The relative emphasis that topics received on the test was judged to be acceptably close to the emphasis these topics received in teacher education curricula. Differences in emphasis reflected an overrepresentation of topics on the test. Ninety-two percent of the panelists judged that the test closely paralleled, or did not differ significantly from, the content of Louisiana teacher education programs.

Listening Test

Ninety-eight percent of the items were rated as content appropriate. Two topics were cited as omitted from the test: analysis and synthesis of oral communication, and stimulus-response questions for communication. The relative emphasis was .36, indicating that less than

50 percent of the panelists judged that topics were given the same emphasis on the test that they received in teacher education programs. Differences in emphasis were generally cases in which topics were overrepresented on the test. Forty percent of the panelists rated the test as closely paralleling, or not differing significantly from, Louisiana teacher education programs.

Writing Test

As noted earlier, Content Review was limited to the objective section of this test. Ninety-six percent of the items were judged to be content appropriate. One topic was noted by two judges as omitted: the spelling of frequently confused words (to-too, threw-through). The relative emphasis value was an acceptable .50. When the emphasis given topics on the test differed from that given in the curricula, topics were overrepresented on the test. In judging overall similarity, 64 percent of the panelists evaluated the test as matching the content of Louisiana teacher education programs.

KNOWLEDGE ESTIMATION

The Knowledge Estimation Panel members evaluated the difficulty and importance of each item in the eight Core Battery Tests. They then estimated the proportion of items that the minimally knowledgeable teacher candidate would answer correctly. The method used for evaluating the essay exercise of the Writing Test was somewhat different, since the essay was not composed of individual items. For this exercise, panelists reviewed a number of essays that had been written by Louisiana teacher candidates attempting the NTE Core Battery, and classified each as acceptable, unacceptable, or of borderline acceptability.

The proportion of items at each level of difficulty and importance that the minimally knowledgeable teacher candidate could be expected to answer correctly was then combined, for each test, using only those items judged to be appropriate by the Content Review Panel. This procedure ensured that Louisiana Teacher candidates would not be penalized for content they had not had the opportunity to learn. However, teacher candidates could still earn points toward their total score on a Core Battery module by correctly answering those items that had been rated as content inappropriate by the Content Review Panel.

The resulting scores that could be expected of the minimally knowledgeable teacher candidate are shown for each module on Table 25 in Chapter IV. Chapter V discusses the impact of these scores on Louisiana's supply of teacher candidates. Chapter V also shows the percent of teacher candidates who could be expected to attain each set of minimum scores as well as variations from the scores expressed as standard errors of measurement. These estimates were obtained from the performance of Louisiana teacher candidates on the fall, 1982, and spring, 1983, administrations of the NTE Core Battery.

SUMMARY

The Content Review section of this study found the NTE Core Battery modules, in the judgment of those Louisiana college and university faculty members who evaluated their component tests, to be valid measures of teacher education curricula in the State. The Knowledge Estimation results present scores for each module that in the judges' estimation could be expected of a minimally knowledgeable teacher candidate. Information is also provided about the potential effect of possible qualifying scores on the supply of new teachers for Louisiana. All of this information is presented to the Superintendent of Education and the Blue Ribbon Score Committee to assist them in their deliberations about appropriate performance standards on the NTE Core Battery for certification of beginning teachers in the State.

APPENDIX I-A

NTE BLUE RIBBON SCORE COMMITTEE MEMBERS

NTE BLUE RIBBON SCORE COMMITTEE

Committee Member	Representation
Dr. Robert Alciatore, Dean College of Education University of Southwestern Louisiana	Dean Board of Trustees
Dr. Louis Barilleaux, Dean University College Tulane University	Dean Private University
Dr. E. Grady Bogue, Chancellor Louisiana State University Shreveport	Chancellor Board of Supervisors
Ms. Mary Burns	Teacher Teachers' Organization
Mrs. Fran Bussie	American Federation of Labor-Council of Industrial Organization
Ms. Margaret Carmouche	Private Citizen
Mr. Charles Castille	Private Citizen
Dr. John Dempsey Leeville High School	Teacher
Ms. Jackie Ducote Executive Director Louisiana Association of Business and Industry	Louisiana Association of and Industry
Ms. Rosemary Guillory	Teacher
Dr. Elton C. Harrison Vice President for Administration and Placement Dillard University	Private University
Mr. Ben Jeffers	Private Citizen
Dr. Burnett Joiner, Dean Division of Education Grambling State University	Dean Board of Trustees
Mr. Sam Jones President Young Democrates of Louisiana	Young Democrats of Louisiana
Ms. Brenda S. Jordan	Teacher Teachers' Organization
Dr. Tom Kelly, Chairman Louisiana College	Private University

NTE BLUE RIBBON COMMITTEE

Committee Member	Representation
Mr. Nat LaCour President United Teachers of New Orleans	Teachers' Organization
Mr. Elton LaGasse School Board Memeber Jefferson Parish	Local School Board
Dr. Claire Landry	Principal State Board of Elementary and Secondary Education
Mr. Walter Lee Superintendent Caddo Parish	Superintendent of Schools
Dr. Wesley McClure Southern University Baton Rouge	Vice Chancellor Academic Affairs
Mr. James Prescott Executive Secretary Louisiana School Boards' Association	Lousiana School Boards Association
Dr. Marilyn Ray, Dean Southern University New Orleans	Dean Southern University Board of Supervisors
Ms. Rupert Richardson President, NAACP	National Association for the Advancement of Colored People
Mr. Warren Sevin Principal Terrebonne High School	Principal
Ms. Lorraine Slack President, Louisiana Association of Educators	Teacher Teachers' Organization
Dr. Charles W. Smith, Dean College of Education Louisiana State University (Committee Chairman)	Dean LSU-Board of Supervisors
Dr. Larry Tremblay Staff Board of Regents	Board of Regents
Dr. Hulen B. Williams, Dean College of Chemistry and Physics Louisiana State University Baton Rouge	Dean LSU-Board of Supervisors

NTE BLUE RIBBON COMMITTEE

Committee Member	Representation
Dr. Marvin L. Yates, Dean Junior Division Southern University Baton Rouge	Dean Southern University Board of Supervisors
Dr. Ronald Zaccari, Dean College of Education Southeastern Louisiana University Hammond	Dean

APPENDIX I-B

NTE CONTENT REVIEW AND KNOWLEDGE ESTIMATION PANEL MEMBERS

NTE CONTENT REVIEW AND KNOWLEDGE ESTIMATION PANEL MEMBERS

NAME	COLLEGE
Abbott, Jane	Louisiana State University
Ambrose, Margaret	Southern University
Anderson, James	Louisiana College
Armstrong, Clifford	Southern University
Barnes, Arthur D.	Louisiana State University
Barnhill, Viron L.	University of New Orleans
Barnitz, John	University of New Orleans
Bass, Carrol	Southeastern Louisiana University
Beisenherz, Paul C.	University of New Orleans
Bidner, Sara	Southeastern Louisiana University
Birdsong, Theda	Northeast Louisiana University
Bitner, Joe	Southeastern Louisiana University
Black, Joe	Louisiana College
Blankenstein, Mark	Southern University
Blanton, Linda L.	University of New Orleans
Bodet, Gerald P.	University of New Orleans
Bollman, Glen S.	Louisiana State University
Boudreaux, David	Nicholls State University
Brewer, Kay	Louisiana Tech University
Broussard , Mercedes	Southern University
Broussard, Rolland L.	University of Southwestern Louisiana
Brown, Hugh S.	McNeese State University
Brown, James	McNeese University
Bryant, Bill	Northwestern State University
Butler, Robert W.	University of Southwestern Louisiana
Butler, Walter	Southeastern Louisiana University
Byrne, C. M.	McNeese State University
Carlton, Virginia	Centenary College
Carner, Gilbert C.	University of Southwestern Louisiana
Carpenter, Dana	Southern University
Cartledge, Frank K.	Louisiana State University
Casler, Burtis G.	Louisiana State University
Casperon, Luvonia	Louisiana State University
Chachere, Ernest G.	University of New Orleans
Chadick, Stanley	Northwestern State University
Chauvin, Jane C.	St. Mary's Dominican College
Clarke, Wilbur B.	Southern University
Clement, Robert J.	Nicholls State University
Cobb, Thelma	Southern University
Coumes, John V.	Southeastern Louisiana University
Crocker, Bernard W.	University of Southwestern Louisiana
Dakin, Matt E.	University of Southwestern Louisiana
Daspit, Robert	Northwestern State University
Davis, Lawrence H.	Southeastern Louisiana University
Deamer, Thelma L.	Southern University
Deconge, Lovenia	Southern University
DeLatte, Carolyn E.	McNeese State University
Dennis, John H. Jr.	Nicholls State University
Detmers, William R.	Louisiana State University
Dillard, J. L.	Northwestern State University
Dobie, Ann B.	University of Southwestern Louisiana
Dobkins, David H.	University of Southwestern Louisiana
Doney, Hugh H.	Northeast Louisiana University

NTE CONTENT REVIEW AND KNOWLEDGE ESTIMATION PANEL MEMBERS

NAME	COLLEGE
Dorman, Wade	Louisiana State University
Elliot, Richard J.	University of New Orleans
Ellois, Edward Jr.	Southern University
Eversull, LeRoi	Northwestern State University
Ford, Luther L.	Grambling State University
Foshee, Andrew W.	McNeese State University
Foss, Roger V.	Northeast Louisiana University
Fuller, Claude C.	Northeast Louisiana University
Garner, Joseph D.	Centenary College
Gau, Margaret F.	Grambling State University
Gifford, Charles S.	University of New Orleans
Gillie, Nina	Southern University
Greene, Jane F.	Our Lady of Holy Cross College
Handford, Charlene	Louisiana State University
Hanna, Ruth	Louisiana Tech University
Harris, Louise E.	Northeast Louisiana University
Haw, Larry S. Sr.	Nicholls State University
Heleniak, Roman	Southeastern Louisiana University
Hennigan, Thomas	Northwestern State University
Hernandez, Nancy M.	Our Lady of Holy Cross College
Hietter, James G.	Louisiana State University
Hoch, Ivan S.	University of Southwestern Louisiana
Holmes, Billy J.	Nicholls State University
Holmes, Jerry D.	Northeast Louisiana University
Holmes, Lawrence H. Jr.	Southeastern Louisiana University
Hopkins, Mary F.	Louisiana State University
Houston, Jacqueline	Dillard University
Hunt, Jean M.	Grambling State University
Iles, Bill	McNeese State University
Jennings, T. M.	Grambling State University
Jones, Barbara C.	Northeast Louisiana University
Jones, Millard T.	McNeese State University
Jones, Robert L.	University of Southwestern Louisiana
Jones, William V.	Louisiana State University
Jordahl, Robert A.	McNeese State University
Keisler, James E.	Louisiana State University
Kindlen, D.	Louisiana State University-Baton Rouge
King, Jean A.	Tulane University
King, Viola D.	Southern University
Landry, David	Nicholls State University
Lanoux, Sigred	University of Southwestern Louisiana
Larue, Madeleine	St. Mary's Dominican College
Lassiter, Darnell T.	Southern University
Lee, Kil S.	University of New Orleans
Leslie, J. P.	Nicholls State University
Lewis, John C.	Northeast Louisiana University
Lin, San-Su	Southern University
Loy, Barbara	St. Mary's Dominican College
Lumpkins, Bob	Northwestern State University
Lyle, Marguerite R.	University of Southwestern Louisiana
Lyons, Don T.	McNeese State University
MacCurdy, Carol A.	University of Southwestern Louisiana
Maddox, Glenda	Louisiana Tech University

NTE CONTENT REVIEW AND KNOWLEDGE ESTIMATION PANEL MEMBERS

NAME	COLLEGE
McCready, Michael A.	Louisiana Tech University
Metzger, Devon J.	University of New Orleans
Middleton, David	Nicholls State University
Millen, Earl T.	Northeast Louisiana University
Miller, Jerry L.	Louisiana Tech University
Mills, Johnie R.	Grambling State University
Moser, Ernest R.	Northeast Louisiana University
Mulliner, Gladys W.	Southern University
Myrick, Susan	Louisiana College
Nassar, Toffee Jr.	McNeese State University
Nelson, Donald	Louisiana College
Newman, Dorothy	Southern University
Norris, H. E.	Louisiana State University
Odom, Thomas W.	Grambling State University
Ohlendorf, George W.	Louisiana State University
Patterson, Thomas J.	Southern University
Patton, John H.	Louisiana State University
Penfield, Elizabeth F.	University of New Orleans
Phillips, Jenny D.	Grambling State University
Piliawsky, Monte	Dillard University
Pitchford, Henry G.	University of Southwestern Louisiana
Pleasant, John R. Jr.	Southeastern Louisiana University
Portis, Theodore R.	Grambling State University
Presse, Norman J.	Nicholls State University
Prince, Kay R.	Louisiana Tech University
Probst,, C. J. Jr.	University of New Orleans
Pullig, Maurice	McNeese State University
Rankin, Elizabeth D.	Louisiana State University
Rathle, Pierre A.	Nicholls State University
Rausch, John R.	Louisiana State University
Reddix, Roscoe C.	Southern University
Redford, Debbie J.	University of Southwestern Louisiana
Reed, James F.	McNeese State University
Rhodes, Donald	Southeasten Louisiana University
Riehl, Robert	Nicholls State University
Riley, Mignon	Northeast Louisiana University
Robert, William	Northwestern State University
Robertson, Patricia R.	University of Southwestern Louisiana
Robinson, Evelyn	Southeastern Louisiana University
Rogers, Stearns	McNeese State University
Ryder, William	Southern University
Saacks, Marguerite E.	Loyola University
Sabrio, David	Louisiana State University
Sanders, Velora S.	Grambling State University
Sauls, Charles W.	Louisiana State University
Schexnider, Ray	Northwestern State University
Scott, Ann M.	University of Southwestern Louisiana
Self, Robert P.	Nicholls State University
Shane, James C.	Grambling State University
Shaw, Ralph W. Jr.	Southeastern Louisiana University
Sheumaker, Charles R.	University of Southwestern Louisiana
Simms, Earline M.	Grambling State University
Simms-Brown, Ruby	Southern University

NTE CONTENT REVIEW AND KNOWLEDGE ESTIMATION PANEL MEMBERS

NAME	COLLEGE
Simpson, William	Louisiana College
Skinner, William R.	McNeese State University
Smalley, Alfred E.	Tulane University
Smith, Dorothy P.	University of New Orleans
Smith, Fred M.	Louisiana State University
Smith, Jackson	Northeast Louisiana University
Spikes, Dolores R.	Southern University
Swetman, Glenn	Nicholls State University
Taylor, Willene P.	Southern University
Teague, Anna D.	Southeastern Louisiana University
Thames, Mary L.	University of New Orleans
Tharpe, Edith M.	Baptist Christian College
Tully, Anita	Nicholls State University
Wade, Luther	Southeastern Louisiana University
Walker, Barbara C.	Southeastern Louisiana University
Watson, Cresap	University of New Orleans
Webert, Henry	Nicholls State University
Wells, Dorothy B.	Southern University
Whitfield, George	Southern University
Whittaker, Leon	Grambling State University
Whittington, Curtis C.	McNeese State University
Williams, Allen	Grambling State University
Wiltz, Carroll	Dillard University
Wooten, Carl	University of Southwestern Louisiana
Young, Henry	Southern University
Young, John C.	McNeese State University

APPENDIX I-C

LOUISIANA STATE DEPARTMENT OF EDUCATION NTE TASK FORCE MEMBERS

LOUISIANA STATE DEPARTMENT OF EDUCATION NTE TASK FORCE

Dr. Helen Brown, Director, Bureau of Curriculum, Inservice, and Staff Development

Mr. Robert Crew, Director, Bureau of Higher Education and Teacher Certification

Ms. Carol Falkowski, Administrative Officer, Bureau of Research

Dr. Robert Garvue (Task Force Chairman), Assistant Director, Bureau of Research

Mr. Michael Glisson, Assistant Director, Bureau of Higher Education and Teacher Certification

Dr. Lee Hoffman, Section Chief, Bureau of Evaluation

Dr. Jacqueline Lewis, former Director, Bureau of Higher Education and Teacher Certification

Dr. Michael McGuire (Chief Administrator for Task Force), Administrative Officer, Bureau of Management Information Systems

Dr. Craig Mills, Administrative Officer, Bureau of Accountability

Mr. J. Frank Norris, Director, Bureau of Materials of Instruction and Textbooks

Dr. Hugh Peck, Associate Superintendent for Research and Development

Dr. Janella Rachal, Administrative Officer, Bureau of Evaluation

Dr. David Ransen, Bureau of Management Information Systems

Dr. Charles Teddlie, Administrative Officer, Bureau of Research

Mr. Joseph Williams, Jr., Director, Bureau of Accountability

APPENDIX III-A

OVERVIEW OF CONTENT REVIEW PANEL TASKS

OVERVIEW OF TASKS TO BE PERFORMED BY MEMBERS OF CONTENT REVIEW PANELS

The study in which you have been asked to participate is being conducted by the Louisiana State Department of Education (LSDE). The purpose of the study is to evaluate the content of the National Teacher Examinations (NTE) in relation to Louisiana teacher education programs and to estimate the test performance of minimally knowledgeable candidates for certification as teachers in the public schools in Louisiana.

A Content Review Panel has been established for each section of the NTE Core Battery Tests. You have been selected to serve on the Content Review Panel for-- ________________________ ______________________ ____________________.

As a member of the Content Review Panel, you will be asked to perform two tasks:

1. To examine the description of test content that is followed in developing the test and to ascertain whether the curriculum areas covered by the test are congruent with the curriculum areas covered by the program at your institution or other Louisiana institutions with which you are familiar.

2. To examine individual test questions and to judge whether the content of each question would have been included in the teacher education program(s) followed by the students who take the test for certification.

As indicated in the NTE Design Summary, a number of faculty members will be assembled in order to make their judgments. The judgments, however, will be made individually and independently; members of the same panel will not confer as a group, nor will any member be informed of the judgments made by any other individual member. The judgments of all members of a panel will be combined statistically by the LSDE to arrive at a summary judgment for the panel about each question. The summary judgments for the questions also will be combined, and the final results will be published in a report describing the study and its findings or conclusions.

The information in this mailing is intended to help you to prepare for your tasks. If in studying the materials you find that you have questions about the tasks, be sure that they are answered during the initial orientation session at the central meeting site.

The Test Content Description identifies the major groups of topics that are covered by the test and indicates the relative emphasis that is given to each. You will be asked to evaluate the overall congruence between the content of the teacher education curriculum and the content of the test, and to record your evaluation on the Test Content Review Form. In addition, you will be given a set of test questions and asked to make judgments about them and to record your judgments on the Question Review Form. Before you go to the central meeting site, please think carefully about the Test Content Description and the curriculum at your institution. You may want to make preliminary notes while you have access to various sources of information on your own campus and to bring them with you to the meeting site, where you will be asked to complete the Review Forms.

You have been asked to participate in this study because you are familiar with the curriculum at your institution in the field(s) covered by the examination with which you will be working. Before attending the panel session, you may want to draw upon local sources of information regarding the curriculum at your institution, such as your college catalog, specialists in curriculum planning, or other sources available to you. You may also find it helpful to talk with colleagues who have taught specific courses that you have not been called upon to teach.

Your contribution to this study and your qualifications to participate are an important part of the study methodology. In order that the final report of the study's findings and conclusions be as informative as possible to others who may wish to use it, we will ask for your permission to identify you in the final report. Your individual judgments will not be identified.

When you go to the meeting, please take this packet of materials with you.

We very much appreciate your willingness to participate in this important study.

Robert J. Garvue
Chair
NTE Validation Task Force
Louisiana State Department of Education

APPENDIX III-B

QUESTION REVIEW FORM AND INSTRUCTIONS

INSTRUCTIONS

QUESTION REVIEW FORM

Your task is to make judgments about whether or not the content of individual test questions is taught in one or more courses that are part of Louisiana teacher education programs leading to certification.

In making your judgment about each question, consider whether or not the content of the question would have been covered in any of the courses normally taken by students in teacher education programs. In some cases a course that is prerequisite to entering a teacher education program may have been taken in college by some students but in high school by others. When such a course is the one in which the content of a particular question is taught, all students who would have taken the course, whether in college or in high school, should be considered to have had the opportunity to learn the content. You are not to judge whether the students would, in fact, have learned the answer; you are only to judge whether they would have had an opportunity to learn the answer.

As you read each test question and its underlined answer, judge whether at least 90 per cent of the students in the group with which you are concerned would have had an opportunity to learn the answer. If you think that they would have, circle "Yes" on the Question Review Form; if you think that they would not have, circle "No". Before you circle your answer, please make sure that the number that identifies the question on the answer form is the same as the number that identifies the question in the question set.

If you feel that your experience provides you with no basis whatsoever for making a judgment about one of the questions, you may circle "DNK" (for "Do Not Know"). The DNK category is not to be used simply because you have difficulty in deciding whether to answer "Yes" or "No"; you are to make a decision even if it is a difficult one. The DNK category is to be used only when you have no basis for making any judgment.

In making your judgments you are not to be concerned about how many questions you are assigning to the "Yes" category or to the "No" category. It is your responsibility to apply your best judgment in evaluating each question individually.

After you have finished making your judgments about the questions in the entire question set, please look over the questions and your decisions about them to make sure that you are satisfied with the consistency of your judgments.

* * *

The operators who will be keypunching the forms will appreciate your using a Number 2 pencil and erasing carefully when you change your mind so that your final judgment about each question will be clearly indicated.

As you turn to each new test item, please be sure that the number of the question on the page corresponds to the number on the Question Review Form. When you have made judgments about all questions in the question set, check that the number of the last question for which you have recorded a judgment on the form corresponds to the number of the last question in the question set.

QUESTION REVIEW FORM

Professional Knowledge: SECTION 1

______________________________ Judge's Name

A
Form

Question Number	Is Content Taught?	Question Number	Is Content Taught?	Question Number	Is Content Taught?
1.	Yes No DNK	13.	Yes No DNK	25.	Yes No DNK
2.	Yes No DNK	14.	Yes No DNK	26.	Yes No DNK
3.	Yes No DNK	15.	Yes No DNK	27.	Yes No DNK
4.	Yes No DNK	16.	Yes No DNK	28.	Yes No DNK
5.	Yes No DNK	17.	Yes No DNK	29.	Yes No DNK
6.	Yes No DNK	18.	Yes No DNK	30.	Yes No DNK
7.	Yes No DNK	19.	Yes No DNK	31.	Yes No DNK
8.	Yes No DNK	20.	Yes No DNK	32.	Yes No DNK
9.	Yes No DNK	21.	Yes No DNK	33.	Yes No DNK
10.	Yes No DNK	22.	Yes No DNK	34.	Yes No DNK
11.	Yes No DNK	23.	Yes No DNK	35.	Yes No DNK
12.	Yes No DNK	24.	Yes No DNK		

APPENDIX III-C

TEST CONTENT DESCRIPTIONS,[1] TEST CONTENT REVIEW FORMS, AND INSTRUCTIONS

---

[1] Permission to reproduce copyrighted material granted by Educational Testing Service

NTE Core Battery Test Content Description

Communication Skills: READING

Topic	Approximate % of Test
I. **Comprehension**	50
II. **Analysis**	35
III. **Evaluation**	15

I. Comprehension — 50

The ability to understand accurately and completely the explicit content of a written message

1. Main Idea
2. Detail
   e.g., Definition - word, phrase, etc.
   Supporting Ideas
3. Relationships
   e.g., Sequence
   Cause and Effect
4. Paraphrase/Summary

II. Analysis — 35

The ability to clarify a written message and to understand how it is organized and conveys its message

1. The writer's purpose
2. The writer's assumptions
3. The writer's attitude or tone
4. Implications of the message
   Inferences from the message
5. Fact vs. opinion in the message
6. Organization of the message
7. Use of language in the message
8. Application of elements in the message

III. Evaluation — 15

The ability to make reasoned qualitative judgments about the nature and merits of a written message

1. Emotional or manipulative aspects of the message
2. Strengths and/or weaknesses of the argument
3. Relevance and/or appropriateness of supporting evidence, arguments
4. Relation of the message to the audience and/or to the general universe of the topic

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

NTE Core Battery Test Content Description

Communication Skills: WRITING

	Topic	Approximate % of Test
I.	Usage - including capitalization and punctuation, subject-verb agreement, verb form, pronoun problems, parallelism, diction, idiom, structural problems, and adjective-adverb confusion	55
II.	Sentence Correction - including problems of coherence, word order, economy of statement, appropriateness of diction and choice of idiom, subordination of sentence elements, logical comparison structure, and clarity of modification and pronoun reference	45

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

NTE Core Battery Test Content Description

General Knowledge: LITERATURE AND FINE ARTS

	Topic	Approximate % of Test
I.	Recognizing basic elements and components of works of literature and fine arts	29
II.	Analyzing and interpreting works of literature and fine arts	54
III.	Relating works of literature and fine arts to one another and to their social/historical context	17

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

NTE Core Battery Test Content Description

General Knowledge: MATHEMATICS

	Topic	Approximate % of Test
I.	Has good number sense and understands how numbers behave	20
II.	Understands and uses numbers in an appropriate way to quantify thinking	16
III.	Recognizes and uses mathematical relationships	24
IV.	Understands the mathematical basis of measurement	16
V.	Understands deductive reasoning	12
VI.	Can interpret graphic, symbolic, and verbal material	12

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

NTE Core Battery Test Content Description

General Knowledge: SCIENCE

	Topic	Approximate % of Test
I.	Demonstrates understanding of energy relationships in both living and nonliving contexts	11
II.	Demonstrates understanding of the significant features of living things	11
III.	Demonstrates understanding of the fact that the operation of natural processes has resulted in organisms that fill a vast number of ecological niches and that these organisms are usually classified on a structural basis into a small number of categories, which facilitate the understanding and study of the organisms	11
IV.	Demonstrates understanding of the relationships between living organisms, particularly humans, and the environment	11
V.	Demonstrates understanding of the fact that Earth is both a part of the Universe and a body that has special characteristics	11
VI.	Demonstrates understanding that all matter is composed of atoms, that atoms are divisible, and that atoms undergo combinations	11
VII.	Demonstrates understanding of the forces that act on units of matter	11
VIII.	Demonstrates understanding of the methods of science: the kinds of reasoning and the organization of information that have contributed to the development of science	11
IX.	Demonstrates understanding of the role of science in securing and maintaining important human values	11

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

NTE Core Battery Test Content Description

General Knowledge: SOCIAL STUDIES

	Topic	Approximate % of Test
I.	Understanding the forces which have influenced the evolution and current state of human culture and institutions	25
II.	Understanding the behavior of individuals, of small groups, and of social institutions and the interrelationships among individuals, groups, and social institutions	25
III.	Recognizing both the universal features of world culture and history, and the basic differences among cultural and national units	25
IV.	Possessing the essential tools and the balanced perspective to analyze and make informed judgments about society	25

The topic areas above will be related to the following specific subject matter: major U.S. historical and cultural events and movements; political institutions and political values; prominent characteristics of societies and cultures (e.g., patterns of social change, political organizations, political values); relationship between culture and individual (e.g., processes and patterns of prejudice, stereotyping, and discrimination); economic concepts and processes; geographical features and characteristics of human settlement and culture; and methodologies, methodological tools, and resources of social sciences.

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

NTE Core Battery Test Content Description

PROFESSIONAL KNOWLEDGE

	Topic	Approximate % of Test
I.	Planning objectives, diagnosing needs, identifying resources, and designing instruction	24
II.	Implementing conditions that facilitate learning and instructional design	25
III.	Evaluating student achievement and instructional effectiveness and using evaluation data to refine instruction	17
IV.	Recognizing students' constitutional rights and state, federal, and judicial policy, and their implications for classroom practice	9
V.	Recognizing extra-classroom influences on teachers and students, including school policy, community expectations, the media, and children's developmental patterns	11
VI.	Demonstrating knowledge of the teaching profession and of professional teaching behaviors	14

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

NTE Core Battery Test Content Description

Communication Skills: LISTENING

	Topic	Approximate % of Test
I.	Basic Comprehension of Message (includes paraphrasing message, understanding connotations of words, and summarizing major idea)	37
II.	Analysis of Message (includes identifying assumptions, drawing inferences, recognizing implications, and identifying speaker's tone)	30
III.	Evaluation of Message (includes identifying and evaluating logical structure, assessing appropriateness and effectiveness of supporting material, and evaluating effect of speaker's tone on an audience)	18
IV.	Feedback-Response (includes identifying appropriate responses to questions or dialogues)	15

Copyright © 1982 by Educational Testing Service. All rights reserved.

Educational Testing Service is an Equal Opportunity Employer.

INSTRUCTIONS

CONTENT REVIEW FORM

The following chart lists the major content topics that are covered in this NTE test, along with the relative weight given to each topic. Compare the relative emphasis of the topics listed to the relative emphasis of these topics within the broad curriculum outline of the teacher education sequence at your institution.

Circle the "S" next to each topic that is given about the same emphasis in the teacher education curriculum at your institution as in the test. Disregard small percentage differences, i.e., differences of 5% or less.

Circle the "M" next to each topic that is given more emphasis in your institution's teacher education curriculum than in the test. Mark this column only if in your opinion the difference is greater than 5%.

Circle the "L" next to each topic that is given less emphasis in your institution's teacher education curriculum than in the test. Mark this column only if in your opinion the difference is smaller than 5%.

CONTENT REVIEW FORM

PROFESSIONAL KNOWLEDGE

____

Judge's Name

Content Topic* PROFESSIONAL KNOWLEDGE	% of test	Emphasis in the Curriculum		
		SAME	MORE	LESS
I	24	S	M	L
II	25	S	M	L
III	17	S	M	L
IV	9	S	M	L
V	11	S	M	L
VI	14	S	M	L

* Refer to the accompanying Test Content Description as you complete this form.

TEST CONTENT OMISSION FORM

____
Judge's Name

Use the space below to list any major content areas in your institution's teacher education sequence that do not appear to be covered in the Test Content Description. Please note that the topics you write in should be of approximately the level of specificity as those listed. Do not list topics that may be subsumed under one of the categories listed - if you are uncertain whether to list a topic here, please refer back to the Test Content Description before doing so.

APPENDIX III-D

TEST CONTENT SUMMARY FORM

TEST CONTENT SUMMARY FORM

____
Judge's Name

Based on the information you have regarding the general content topics covered in this test, select the option below that most closely characterizes your judgment regarding the similarity between this NTE test and the teacher education sequence at your institution. Indicate your answer by placing an X next to the response that you have chosen, then use the space below to add any additional comments you may have.

____ The test content topics parallel the teacher education sequence at our institution very closely.

____ There are some differences between the test content topics and the teacher education sequence at our institution, but these differences do not appear to be appreciable.

____ There appear to be some appreciable differences between the test content topics and the teacher education sequence at our institution.

____ There is little similarity between the test content topics and the teacher education sequence at our institution.

APPENDIX IV-A

ITEM ESTIMATION FORM

ITEM ESTIMATION FORM

Professional Knowledge: SECTION 1

____
Judge's Name

A
Form

Directions for Rating Item Importance

Read each test item and the answer choices which accompany it carefully. As you read the item, think of the importance of the knowledge required to answer the item correctly. In your opinion, does this item test essential knowledge (E), important knowledge (I), or not very important knowledge (N)? Circle the E, I, or N in the column labelled "Item Importance" to reflect your judgment.

Directions for Rating Item Difficulty

As you read each item, think of how the minimally knowledgeable student would perform on this item. In your opinion, would the minimally knowledgeable student find this to be an easy item (E), a moderately difficult item (M), or a hard (H) item? Circle the E, M, or H in the column labelled "Item Difficulty" to reflect your judgment.

(When you have rated all items in both question sets, go on to the Test Estimation Form.)

Item Number	Item Importance			Item Difficulty			Item Number	Item Importance			Item Difficulty		
1.	E	I	N	E	M	H	19.	E	I	N	E	M	H
2.	E	I	N	E	M	H	20.	E	I	N	E	M	H
3.	E	I	N	E	M	H	21.	E	I	N	E	M	H
4.	E	I	N	E	M	H	22.	E	I	N	E	M	H
5.	E	I	N	E	M	H	23.	E	I	N	E	M	H
6.	E	I	N	E	M	H	24.	E	I	N	E	M	H
7.	E	I	N	E	M	H	25.	E	I	N	E	M	H
8.	E	I	N	E	M	H	26.	E	I	N	E	M	H
9.	E	I	N	E	M	H	27.	E	I	N	E	M	H
10.	E	I	N	E	M	H	28.	E	I	N	E	M	H
11.	E	I	N	E	M	H	29.	E	I	N	E	M	H
12.	E	I	N	E	M	H	30.	E	I	N	E	M	H
13.	E	I	N	E	M	H	31.	E	I	N	E	M	H
14.	E	I	N	E	M	H	32.	E	I	N	E	M	H
15.	E	I	N	E	M	H	33.	E	I	N	E	M	H
16.	E	I	N	E	M	H	34.	E	I	N	E	M	H
17.	E	I	N	E	M	H	35.	E	I	N	E	M	H
18.	E	I	N	E	M	H							

APPENDIX IV-B

OVERVIEW OF KNOWLEDGE ESTIMATIONS PANEL TASKS

OVERVIEW OF TASKS TO BE PERFORMED BY MEMBERS OF KNOWLEDGE ESTIMATION PANELS

The study in which you have been asked to participate is being conducted by the Louisiana State Department of Education (LSDE). The purpose of the study is to evaluate the content of the National Teacher Examinations (NTE) in relation to Louisiana teacher education programs and to estimate the test performance of minimally knowledgeable candidates for certification as teachers in the public schools in Louisiana.

A Knowledge Estimation Panel has been established for each section of the NTE Core Battery Tests. You have been selected to serve on the Knowledge Estimation Panel for -- ______________________ ______________________.

As a member of the Knowledge Estimation Panel, you will be asked to make judgments about the difficulty of individual test questions for persons who have the minimum amount of knowledge necessary to complete the teacher education program required for certification in Louisiana, and to teach effectively. You will also be asked to make judgments about the importance of the knowledge required to answer each test question correctly, and to estimate the number of test items which fall into each category of difficulty and importance. Your judgments will be combined with judgments made by other faculty members to derive an estimate of the probable test performance of this group of persons.

As indicated in the NTE Design Summary, a number of faculty members will be assembled in order to make their judgments. There will be two meetings of the Knowledge Estimation Panel: a Training Session on October 14, 1982, to provide a detailed group orientation to and exercise in the Knowledge Estimation procedures; and the actual Rating Session on October 15, 1982. At this latter time, judgments will be made individually and independently; members of the same panel will not confer as a group, nor will any member be informed of the judgments made by any other individual member. The judgments of all members of a panel will be combined statistically by the LSDE to arrive at a summary judgment for the panel about each test. The summary judgments for the tests also will be combined, and the final results will be published in a report describing the study and its findings or conclusions.

The information in this mailing is intended to help you to prepare for your tasks. If in studying the materials you find that you have questions about the tasks, be sure that they are answered during the Training Session on October 14. The Test Content Description identifies the major groups of topics that are covered by the test and indicates the relative emphasis that is given to each. It will serve to familiarize you with the general content of the test before you see the test questions themselves.

You have been asked to participate in this study because you are familiar with the curriculum at your institution in the field(s) covered by the examination with which you will be working. Before attending the panel session you may want to draw upon local sources of information regarding the curriculum at your institution, such as your college catalog, specialists in curriculum planning, or other sources available to you. You may also find it helpful to talk with colleagues who have taught specific courses that you have not been called upon to teach.

Your contribution to this study and your qualifications to participate are an important part of the study methodology. In order that the final report of the study's findings and conclusions be as informative as possible to others who may wish to use it, we will ask for your permission to identify you in the final report. Your individual judgments will not be identified.

When you go to the meeting, please take this packet of materials with you.

We very much appreciate your willingness to participate in this important study.

Robert J. Garvue
Chair
NTE Validation Task Force
Louisiana State Department of Education

APPENDIX IV-C

DESCRIPTION OF THE MINIMALLY KNOWLEDGEABLE TEACHER CANDIDATE

Description of the Minimally Knowledgeable Teacher Candidate

For the purpose of this study, use as a frame of reference the minimal amount of academic knowledge needed to a) complete the college program required for certification in the State, and b) teach effectively. An assessment of minimal knowledge is a task that every faculty member performs, at least in part, every time he or she writes and grades examinations for students enrolled in college courses or evaluates a student teacher's performance. The dividing line between a minimally passing and a failing grade must be established by a faculty member in designing an examination so as not to set the level of difficulty of the questions so high that it excludes the minimally knowledgeable student from demonstrating the level of knowledge he or she has. That dividing line must be assessed again in grading examinations and, because the delivery of a failing grade has such important consequences, most conscientious faculty members pay great attention to their conception of what the minimally knowledgeable student should be able to do to achieve a passing grade on their examinations. You will now be asked to draw upon this experience in applying your conception of the minimally knowledgeable student to test questions written for the NTE.

APPENDIX IV-D

INSTRUCTIONS FOR KNOWLEDGE FORMS AND KNOWLEDGE ESTIMATION TRAINING EXERCISE

INSTRUCTIONS

KNOWLEDGE ESTIMATION FORMS

Your task is to make judgments about the difficulty and importance of individual test questions for minimally knowledgeable persons in all teacher education fields. You will be asked to draw upon your experience to construct a hypothetical group of persons, each of whom, in your judgment, has the minimum amount of academic knowledge to complete the teacher education program required for certification in Louisiana, and has the minimum amount of knowledge to teach effectively. The standards that you apply should be appropriate to both criteria, that is, completing the academic program and then using the knowledge gained from the academic program to teach. In drawing upon your experience, you will probably find three types of persons for whom different levels of achievement would represent minimum knowledge: (1) those who will pursue nonteaching careers after graduation; (2) those who will pursue elementary or secondary teaching careers; and (3) those who will pursue graduate study. This study is concerned only with persons who will pursue elementary or secondary teaching careers, and the assessments that you make with respect to minimally knowledgeable persons should be made with only this category of persons in mind.

Your judgments about the test questions are to be made with reference to your conception of a group of minimally knowledgeable students, as described in the preceding paragraph. As you read each test question and its underlined answer, think of this group. Then judge whether the minimally knowledgeable student would find each item to be easy(E), moderately difficult(M), or hard(H). Circle the E, M, or H in the column labelled "Item Difficulty" on the Item Estimation Form to reflect your judgment. Now, think of the importance of the knowledge required to answer the item correctly. In your opinion, does this item test essential knowledge (E), important knowledge (I), or not very important knowledge (N)? Circle the E, I, or N in the column labelled "Item Importance" on the Item Estimation Form to reflect your judgment.

If you feel that your experience provides you with no basis whatsover for making a judgment about one of the questions, you may choose not to rate that item. Do not skip items simply because you have difficulty in making a judgment; you are to make a decision even if it is a difficult one. You would choose not to rate an item only when you have no basis for making any judgment.

In making your judgments you are not to be concerned about how many questions you are assigning to the various categories. It is your responsibility to apply your best judgment in evaluating each question individually.

After you have finished recording your judgments about the questions in the entire question set, please look back to make sure that you have applied your standards consistently throughout the entire set.

The operators who will be keypunching the forms will appreciate your using a Number 2 pencil and erasing carefully when you change your mind so that your final judgment about each question will be clearly indicated.

As you turn to each new test item, please be sure that the number of the question on the page corresponds to the number on the Item Estimation Form. When you have made judgments about all the questions in the question set, check that the number of the last question for which you have recorded a judgment on the form corresponds to the number of the last question in the question set.

After you have rated all items in both question sets, you will be asked to complete a second form. Review the decisions you have made about the ratings of item difficulty and importance. Consider the items you have rated as easy and essential. In your opinion, what percent of these items would be answered correctly by the minimally knowledgeable student? In other words, if there were 100 easy/essential items, how many would the minimally knowledgeable student be able to answer correctly? Make a similar judgment for each of the nine combinations of difficulty and importance. Record your judgments in the spaces provided on the Test Estimation Form.

* * *

KNOWLEDGE ESTIMATION TRAINING EXERCISE

Directions for Rating Item Difficulty

Read each item and the answer choices which accompany it carefully. As you read the item, think of how the minimally knowledgeable student would perform on this item. In your opinion, would the minimally knowledgeable student find this to be an east item (E), a moderately difficult item (M), or a hard item (H)? Circle the E, M, or H in the column labelled "Item Difficulty" to reflect your judgment.

Directions for Rating Item Importance

As you read each item, think of the importance of the knowledge required to answer the item correctly. In your opinion, does this item test essential knowledge (E), important knowledge (I), or not very important knowledge (N)? Circle the E, I, or N in the column labelled "Item Importance" to reflect your judgment.

Item Number	Item Difficulty	Item Importance
1	E M H	E I N
2	E M H	E I N
3	E M H	E I N
4	E M H	E I N
5	E M H	E I N
6	E M H	E I N
7	E M H	E I N
8	E M H	E I N
9	E M H	E I N
10	E M H	E I N

APPENDIX IV-E

TEST ESTIMATION FORM

TEST ESTIMATION FORM

Judge's Name ____________________ Test Content ____________________

Directions

When you have finished rating both question sets for a test, complete this form for that test. Be sure to complete the blanks at the top of this form. Then, review the decisions you have made about the ratings of item difficulty and importance. Consider the items you have rated as easy and essential. In your opinion, what percent of these items would be answered correctly by the minimally knowledgeable student? Make a similar judgment for each of the nine combinations of difficulty and importance. Record your judgments in the spaces below.

DIFFICULTY	IMPORTANCE	% OF ITEMS
EASY	ESSENTIAL	
EASY	IMPORTANT	
EASY	NOT VERY IMPORTANT	
MODERATE	ESSENTIAL	
MODERATE	IMPORTANT	
MODERATE	NOT VERY IMPORTANT	
HARD	ESSENTIAL	
HARD	IMPORTANT	
HARD	NOT VERY IMPORTANT	

APPENDIX IV-F

KNOWLEDGE ESTIMATION TRAINING WORKSHEET

Knowledge Estimation Training Worksheet

1. Enter the number of items you rated as Easy/Essential, Easy/Important, etc. into each cell of the matrix.

2. Multiply each of these figures by 10, and enter this new number on the next line in each cell of the matrix. The purpose of this multiplication is to increase artificially the number of items in each cell while maintaining the same proportions as the original rated items.

3. Enter the percentages of items at each level of difficulty and importance which the minimally knowledgeable student should answer correctly into each cell of the matrix, on the line marked "%". These are the percentages you entered on the Test Estimation Form.

4. Multiply this percentage by the number in the "Items X 10" space; enter the product in the space marked "Correct". For example, if there are 20 items in the "Items X 10" space and 90 in the "%" space, the number in the "Correct" space would be 20 X .90 = 18. Perform this tabulation for each cell in the matrix.

Item Difficulty

	EASY	MODERATE	HARD
ESSENTIAL	___ Items ___ Items X 10 ___ % ___ CORRECT	___ Items ___ Items X 10 ___ % ___ CORRECT	___ Items ___ Items X 10 ___ % ___ CORRECT
IMPORTANT	___ Items ___ Items X 10 ___ % ___ CORRECT	___ Items ___ Items X 10 ___ % ___ CORRECT	___ Items ___ Items X 10 ___ % ___ CORRECT
NOT VERY IMPORTANT	___ Items ___ Items X 10 ___ % ___ CORRECT	___ Items ___ Items X 10 ___ % ___ CORRECT	___ Items ___ Items X 10 ___ % ___ CORRECT

Item Importance

END OF DOCUMENT